Imperial College London

Underwater Vision

Maxime Bories

A Thesis submitted in fulfilment of requirements for the degree of
Master of Science
Communications and Signal Processing
of Imperial College London

Department of Electrical and Electronic Engineering
Imperial College London
April 1, 2023

Abstract

Images taken from underwater vehicles suffer from the propagation properties of light in murky water. Attenuation, scattering and the use of an artificial light source result in near-lighting, low contrast and colour incoherence. In this thesis we explore different ways to counter those effects with software, first with common image processing algorithms and then with model-based methods. We will introduce a new physics-based approach specifically designed for artificially lighted underwater scenes. This two-steps method yields an estimation of the scene radiance and the scene albedo. Results on various testing images will demonstrate the efficiency of the method.

Acknowledgment

I would like to thank my supervisor Dr T-K Kim for making the project proposal available and trusting me with the project. Akis Tsiotsios have been here every steps of the project to give me the best of help, share his knowledge and guide me through the work of research. I am deeply grateful for the support and endless show of love from my friends and family back home and in London. Special thanks go to Alex for making coffee breaks during the project much more interesting, and Junior for the holiday break.

Contents

List of Figures

List of Tables

Abbreviations

UUV: Unmanned Underwater Vehicles

ROV: Remotely Operated Vehicles

AUV: Autonomous Underwater Vehicles

LOS: Line-Of-Sight

CLAHE: Contrast Limited Adaptive Histogram Equalization

SfM: Structure from Motion

SIFT: Scale-Invariant Feature Transform

BCC: Brightness Constancy Constraint

Chapter 1

Introduction

Exploration of the underwater world for scientific purposes mainly rely on unmanned underwater vehicles. These underwater drones can be either Remotely Operated Vehicles or Autonomous Underwater Vehicles and serve various purposes such as ocean floor surveying, oceanic cables inspection or underwater mines detection.

Cameras mounted on underwater vehicles produce sequences of underwater images of poor quality due to the behaviour of light in the aquatic environment. Unlike in clear atmospheric photography, light suffers from attenuation and scattering between the light source, the scene and the camera. This interaction of light with the medium results in near-lighting effects, blur, low contrast and color incoherence.
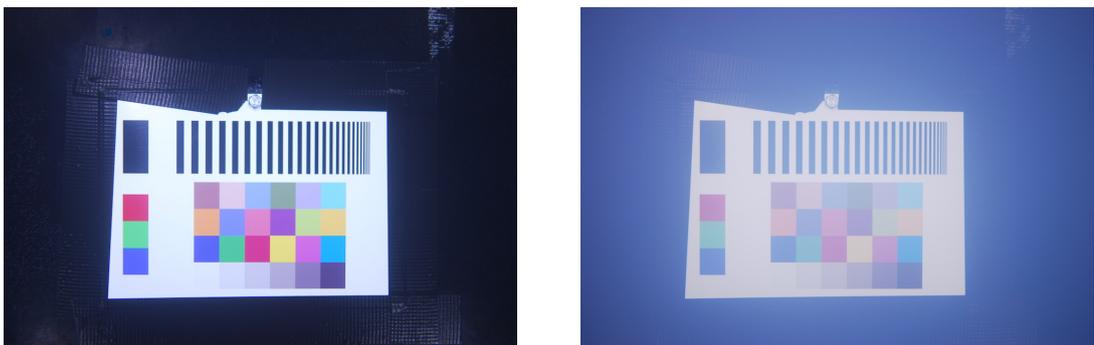
Figure 1.1: Test card imaged in pure water and a water-milk mix simulating the effects of underwater vision.

Whether these images are used for real-time operation of the vehicle or posterior analysis, eliminating those effects is crucial to reveal all the details of the scene.

Ocean floor surveying is a common application of underwater imaging in which sequences of images are taken from above with a moving vehicle. Because of attenuation and scattering, the scene has to be imaged from close, but the study of the floor for marine biology, geology or archeology may necessitate to capture vast areas [2]. A global description of the floor can be obtained by images mosaicing or recovering structure from motion, but those operations can be done only after the effects of underwater vision have been addressed [3] [4].
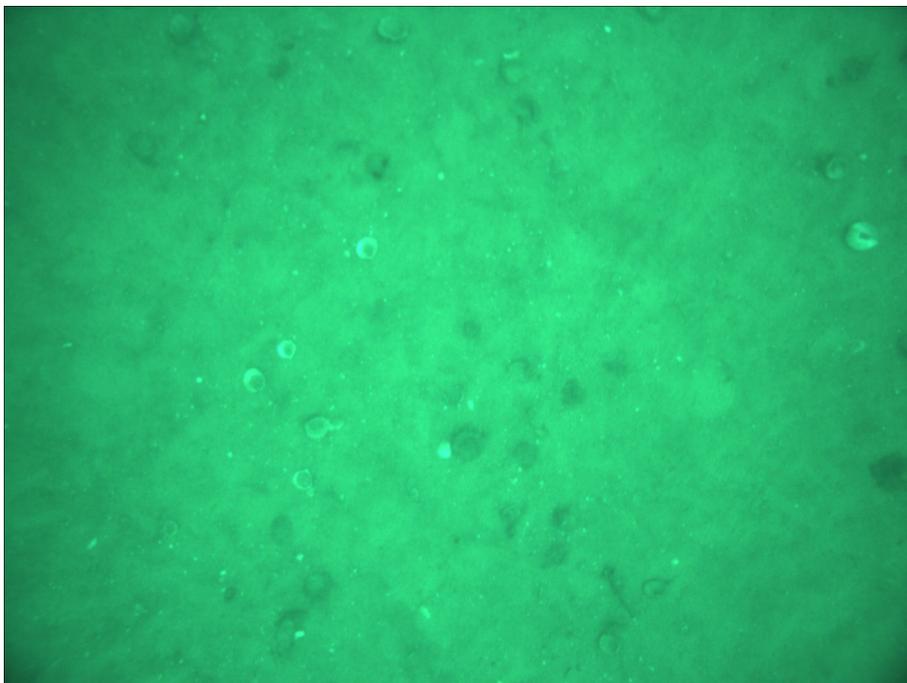
Figure 1.2: Example of ocean floor imaging with green backscatter, low contrast and non-uniform lighting.

Furthermore, we can expect completely autonomous underwater vehicles driven by computer vision to become mainstream in the coming years, but algorithms being currently developed usually assume that the image provided by the camera is the direct scene radiance, hence the need for a pre-processing that eliminates all unwanted effects specific to underwater vision.

1.1 Aquatic Lighting

Attenuation and scattering are both the result of the interaction of light with the various small particles composing the medium along water. Attenuation is the absorption of a constant proportion of incident light energy through a given volume of homogeneous medium that results in heat. Scattering on the other hand, is the change of direction of incident light which spacial distribution is described by a scattering function. Both yield an exponential attenuation of light intensity in the incident direction as a function of the distance between the light source, the scene and the camera. This makes the received image of the scene darker and since the absorption and scattering coefficients are functions of the wavelength the received scene image is not colour-coherent.

Scattering yields another more problematic effect: light coming from the source(s) is scattered back to the camera, creating a haze curtain between the scene and the camera. This additive effect is called backscatter or simply haze in atmospheric photography and drastically diminishes the image contrast. The effects of multiple scattering are more complex and result mainly in blur and modification of the scene illumination [5].

Moreover sunlight is quickly attenuated after a few meters depth, therefore in most cases underwater vehicles need to carry their own artificial light source [6] [7]. This leads to a third unwanted effect: non-uniform lighting of the scene.

To summarise, near-lighting is the result of both using the vehicle light and of the attenuation of light with distance while low contrast is due to attenuation and backscattering.

1.2 State of the Art

The problem of underwater image recovery have been extensively studied in the literature, as well as the similar problem of haze removal in atmospheric images.

Some of theses methods require special hardware installation on the vehicle such as timed lighting, multiple lighting sources, or rotating polarising filters [8] [1] [9]. In this thesis we only describe software methods that can be applied to any footage taken on an underwater vehicle with a single light source.

In 2011 the dark channel prior method was proposed for efficient haze removal of atmospheric images [10]. The uniform illumination of the environment in daylight outdoor photography yields a simpler image formation model so that this method cannot be applied directly to artificially illuminated underwater images. However, the dark channel prior will be used as the starting point in later described methods.

1.3 Proposed Methods

Most of the existing methods in the literature use the same model that is used for haze removal in atmospheric photography [8] [9]. However that model is valid only when we can assume a uniform illumination of the scene [5]. Therefore these methods can only be applied to images taken in the first few tens of meters of depth where the sunlight is not totally absorbed and an artificial lighting is not required. We will introduce a new model that can be used for artificially illuminated scenes.

In this thesis we explore different ways to model and counter these effects in order to recover the scene radiance and albedo. Novel methods will be proposed to counter the effects of back-scattering, attenuation and non-uniform lighting. The efficiency as well as the processing speed of each method will be assessed.

Chapter 2

Non Model-Based Restoration

First we look at classical image processing techniques that can be used to counter the effects of underwater imaging.

2.1 White Balancing

The most obvious unwanted effect of aquatic images is the colour incoherence induced by attenuation and scattering. Depending on the composition of the medium, they usually yields a blue or green illumination and additive backscatter. White balancing allow to compensate the color cast but there is not a universal white balancing algorithm. The most commons are the white patch algorithm, the gray-world assumption and the shades of gray that are simple and fast [11]. Color constancy algorithms try to estimate the illumination characteristics of the scene by statistical computations on the images. Assuming the illumination is uniform its RGB components are estimated and each color channel is simply normalized to the illumination triplet.

2.1.1 Gray-World Assumption

In the gray-world algorithm introduced by Buchsbaum [12], we assume the average reflectance of the scene is gray. Therefore the illumination can by directly estimated by averaging in each channel all the pixels intensities.

2.1.2   White Patch

The white patch algorithm (sometimes called max-RGB) is based on the assumption made by Land [13] that our visual system does color constancy with the brightest parts of the view. When the scene contains a white patch, the illuminant is multiplied by one in each channel at that location directly yielding its value. In this method, each color channel is simply normalized by the maximum values found in the image.

2.1.3   Shades of Gray

The Minkowski p-norm is defined for $p \geq 1$ as:

$$\|x\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{\frac{1}{p}} \tag{2.1}$$

Applied to color channels for $p = 1$, it is equivalent to the Gray-World method with a simple averaging and for $p = \infty$ it becomes the infinity norm equivalent to the White-Patch algorithm. This generalisation introduced by Finlayson and Trezzi is empirically used with $p = 6$ [14].

2.1.4   Performances

These three methods have been implemented in OpenCV and tested on two sets of images: one obtained in a water-milk mix for testing purpose and one from real surveying image taken with an AUV.

The processing time is shown in table 2.1. The fastest method is the white patch algorithm since it only involves finding maximums, while the most computationally expensive is the Shades of Gray algorithm.

The results are shown in figures 2.1 and 2.2. For the test card image the algorithm that performs best is the white patch because the image actually contains white parts. In the case of ocean floor images there is not necessarily a white patch and as shown in figure 2.2 the shade of gray algorithm with $p = 6$ performs better with more neutral tones.

We will therefore chose this algorithm for later methods that contains a white balancing operation.

	Gray-World	White Patch	Shades of Gray
Test Card (1920x1285)	0.507	0.219	0.536
Ocean Floor (1600x1200)	0.370	0.125	0.375

Table 2.1: Processing time of the three white balancing methods in seconds.
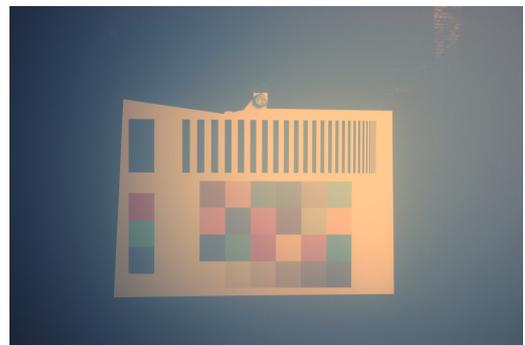
(a) Original image.

(b) Gray-world.
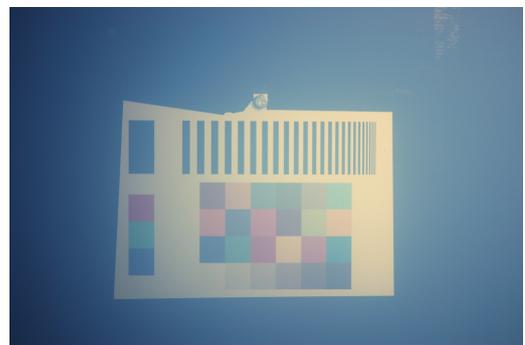
(c) White patch.

(d) Shades of gray with $p = 6$.

Figure 2.1: White balancing results on a test card image.
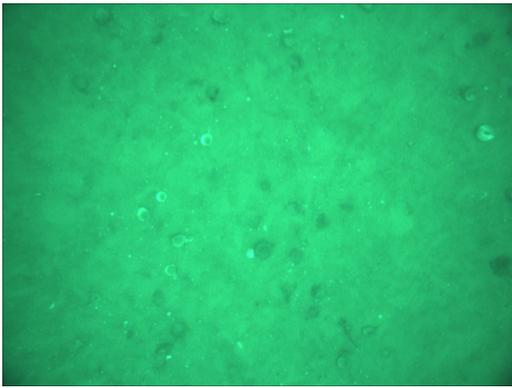
(a) Original image.
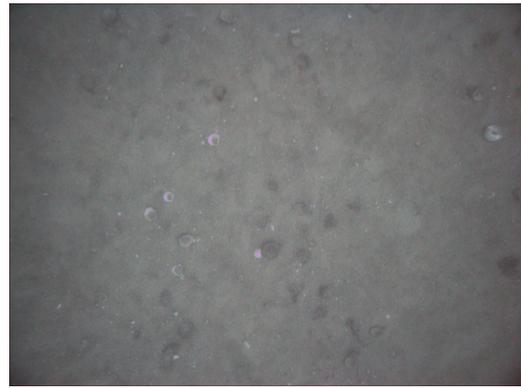
(b) Gray-world.
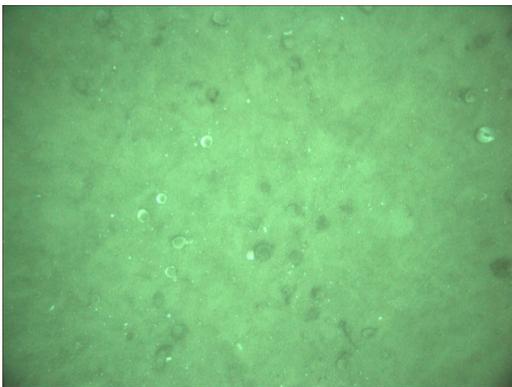
(c) White patch.
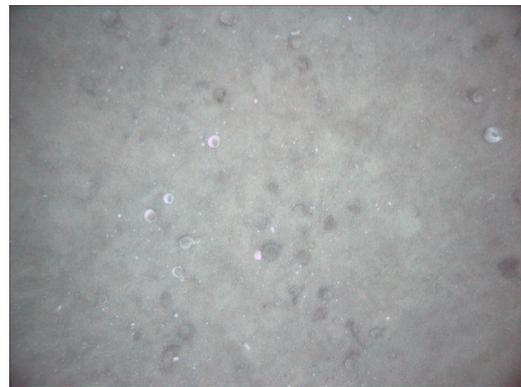
(d) Shades of gray with $p = 6$.

Figure 2.2: White balancing results on ocean floor image.

2.2   Contrasting

After compensating for the color incoherence, we need to address the problem of contrast from the additive backscatter component and the non-uniform lighting. Simple histogram equalization is not an option here because non-uniform lighting yields a non-uniform contrast map. Instead we try the Contrast Limited Adaptive Histogram Equalization algorithm, and the fusion method presented in [15] that have been developed specifically for underwater images.

2.2.1   Contrast Limited Adaptive Histogram Equalization

This algorithm is similar to the local histogram equalization method: for each pixels it equalizes the histogram of the square neighbourhood, but has a contrasting limit that prevents over-amplification of noise [16].

2.2.2   Pyramid Fusion

This method is based on the fusion of two inputs: the white balanced version of the image and a contrast enhanced one. It aims to correct at the same time the problems of color coherence and contrast. The white balanced version is computed with the Shades of Gray algorithm and the contrasted version with the CLAHE algorithm.

For each input four weight maps are computed to highlight all the features of the image: the Laplacian contrast weight, the local contrast weight, the saliency weight and the exposedness weight.
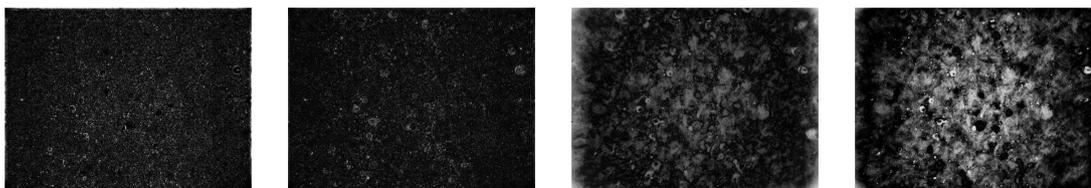
(a) Laplacian contrast.   (b) Local contrast.   (c) Saliency.   (d) Exposedness.

Figure 2.3: Weight maps of the testing image.

The Laplacian contrast weight simply yields global contrast features with a Laplacian filter. The local contrast weight on the other hand is computed by the difference of the pixel value and the Gaussian weighted average of its neighbourhood with a small kernel (for example a 5x5 kernel). The saliency weight is an higher-level map that emphasises objects and is computed here with the algorithm presented in [17]. Finally, the exposedness weight shows how well a pixel is exposed and is computed as the Gaussian distance to the average value:

$$W_E = \exp\left(-\frac{(I - 0.5)^2}{2\sigma^2}\right) \qquad (2.2)$$

Where I is the pixel intensity and $\sigma = 0.25$ the standard deviation of the associated Gaussian distribution.

A naive blending of the inputs and the weight maps yields undesirable halos, therefore a Laplacian pyramid blending is used. The inputs are decomposed into Laplacian pyramids while the weight maps are decomposed into Gaussian pyramid. The fused pyramid is obtained with:

$$R^l = \sum_{k=1}^{K} G^l\{W_k\}L^l\{I_k\} \qquad (2.3)$$

Where l is the pyramid level, $L\{I\}$ is the Laplacian version of the input I and $G\{W\}$ the Gaussian version of the weight map W.

2.2.3   Performances

Again both methods have been implemented in OpenCV and tested on the same sets of image. The two methods greatly improve the contrast, but the fusion method has a slight advantage in both testing images. In the case of the test card, the CLAHE algorithm yields a strong halo which is much less visible with the fusion result as shown in figure 2.4. In the case of the ocean floor, the fusion result looks a little less contrasted but more natural that CLAHE as shown in figure 2.5.

Altogether both methods can be used to solve the contrasting problem, however we still observe vignetting due to the backscatter and the non-uniform lighting of the scene. This problem will be addressed in the next chapter with model-based restoration techniques.

The processing time of both methods is shown in table 2.2. Although the fusion method gives better results, it is to the expense of a processing time five times higher.

	CLAHE	Fusion
Test Card (1920x1285)	0.360	1.821
Ocean Floor (1600x1200)	0.267	1.318

Table 2.2: Processing time of the two contrasting methods in seconds.
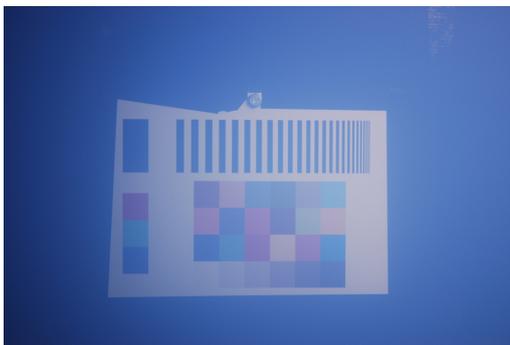
(a) Original Image.

(b) CLAHE.

(c) Fusion.

Figure 2.4: Contrasting results on a test card image.
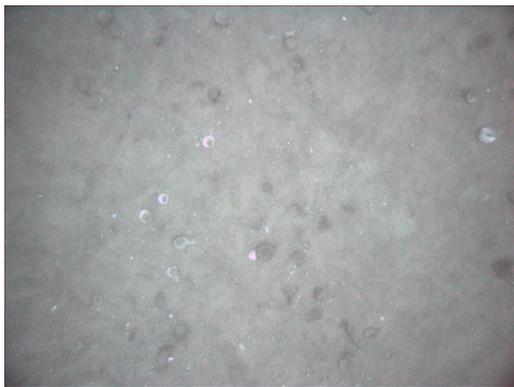
(a) White balanced image.   (b) CLAHE.
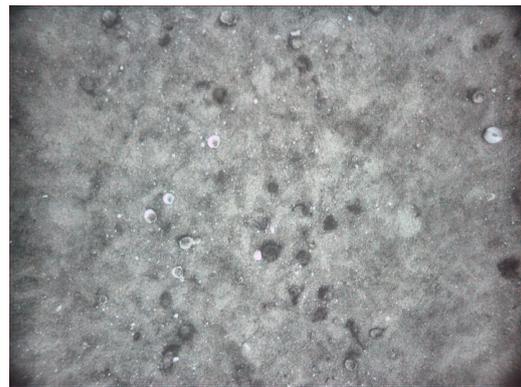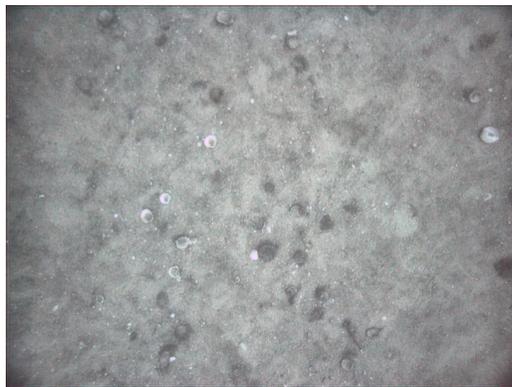
(c) Fusion.

Figure 2.5: Contrasting results on ocean floor image.

Chapter 3

Model-Based Restoration

3.1   Aquatic Lighting

In any transparent medium without interface, incident light always interact with a volume of matter in the same fashion: part of it is transmitted, some is scattered, and another part is absorbed.
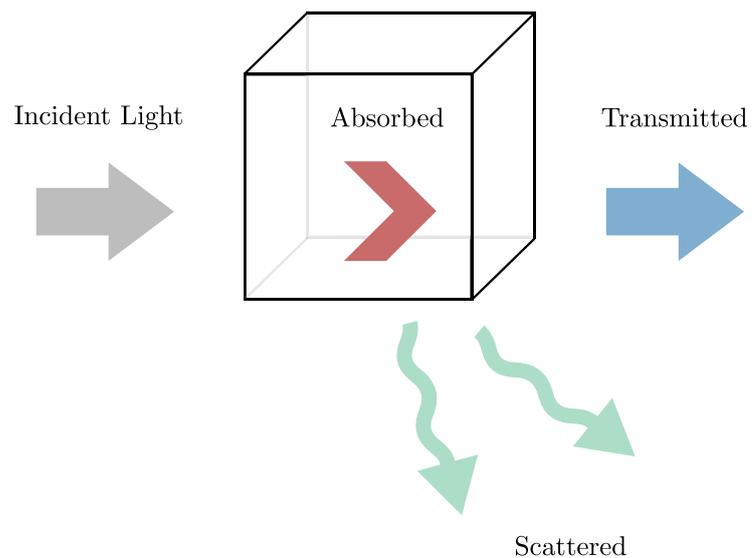
Figure 3.1: Behaviour of light in an elementary volume of material.

The transmitted light will propagate in the same direction as the incident light, while the scattered light will emerge in many directions and intensities that can be described by a scattering function. For example, the Rayleigh scattering function is most often used to describe atmospheric light scattering. Finally, the absorbed part of incident light is converted to heat.

Transmittance T is defined as the fraction of incident electromagnetic power that is transmitted through a given volume after absorption and scattering and is a wavelength dependent function. For a uniform medium, the Beer Lambert law gives the transmittance as a function of the distance travelled by light d:

$$T = e^{-cd} \tag{3.1}$$

The Napierian attenuation coefficient c can be broken down as $c = a + b$ where a and b are the scattering and absorption coefficients respectively.

While the effects of scattering and absorption are present in atmospheric vision - otherwise the sky would always be black- their effect is most often negligible. The particles composing the media are so small that the interaction with incident light is minimal. However in bad weather conditions or polluted environments that contains larger particles, scattering and absorption yields haze or fog that seriously deteriorate the image quality in terms of contrast.

This effect is nearly always present in underwater vision as water isn't pure and larger floating particles are present. All formation models introduced hereafter are therefore of the form:

$$I(x) = D(x) + B(x) \tag{3.2}$$

Where I is the observed intensity, D the direct component is the attenuated scene radiance and B the backscatter component.

3.2   Image Formation Models

If we can assume a uniform illumination of the scene and the medium, that is a light source infinitely distant -such as the sun- the backscatter is only a function of the depth. In that case it can be shown that the direct component $D(x) = J(x)T(x)$ and the backscatter $B = B_\infty(1 - T(x))$ are directly related by the transmittance [5]. The received intensity I is simply observed as:

$$I(x) = J(x)T(x) + B_\infty(1 - T(x)) \tag{3.3}$$

3.2.1   Dark Channel Method

Using this model, the dark channel prior is based on the assumption that in every patches composing the image at least one channel has some pixels close to zero intensity [10]. The dark channel is given by:

$$I_{dark}(x) = \min_{y \in \Omega(x)} \left(\min_{c \in \{r,g,b\}} I^c(y) \right) \tag{3.4}$$

Where $\Omega(x)$ is a square neighbourhood of x. After B_∞ is estimated from the brightest pixels, equation (3.3) is normalized in each channel by B_∞:

$$\frac{I^c(x)}{B_\infty^c} = T(x)\frac{J^c(x)}{B_\infty^c} + (1 - T(x)) \tag{3.5}$$

Applying the dark channel to both sides yields an estimation of the transmission map:

$$T(x) = 1 - \min_{y \in \Omega(x)} \left(\min_{c \in \{r,g,b\}} \frac{I^c(x)}{B_\infty^c} \right) \tag{3.6}$$

The transmission can be eventually refined by soft matting or guided filtering [18], before yielding the scene radiance by:

$$J(x) = \frac{I(x) - B_\infty}{\max(t, t_0)} + B_\infty^c \tag{3.7}$$

Where t_0 is an arbitrary lower bound of the transmission map that avoid dividing noisy patches by near-zero factors.

As long as the initial hypothesis of uniform illumination holds, this method and its derivatives give pleasing results. However in the case of underwater imaging the lighting is uniform only in the first few meters of depth before sunlight is entirely absorbed. In most cases where an underwater vehicle is carrying its own light source, the model itself does not hold because the backscatter component becomes a function of the depth and the illumination along the line-of-sight . Therefore we introduce a similar model that can take into account the non-uniform lighting.

3.2.2   Proposed Model

From the previous model we keep the same direct component $D(x) = J(x)T(x)$ and a linear function for the backscatter $B(x) = \lambda(x)B_\infty$ but with λ a free parameter:

$$I(x) = J(x)T(x) + \lambda(x)B_\infty \tag{3.8}$$

Unlike in the previous model where each pixel has either a more intense direct component or backscatter depending on the value of the transmittance, what we observe with artificially lighted scene is that the backscatter is more intense where the scene radiance is also more intense. The exact correlation between the two component will be given in later advanced modeling.

For now this simple model allow us to separate the direct component from the backscatter with the idea that the backscatter is a linear function of B_∞

In all cases it is better to apply the proposed methods to RAW image files. The first reason is to bypass the non-linear response function found in most camera that leads to physically incoherent intensity values. The second reason is that the dynamic range of

the image is drastically diminished when suppressing the backscatter component. Using RAW files when contrasting the radiance component will avoid over-amplification of the noise.

3.3   Proposed Method for Non-Uniform Depth Images

3.3.1   Estimation of B_∞

Inspired by the dark channel prior, we use the darkest pixels to estimate the backscatter tint B_∞. In underwater imaging the haze is present in every pixels since the camera is surrounded by water. By selecting the 10% darkest pixels in the Euclidean distance sense, (good results are seen between 5% and 30%), we expect mainly near-zero values of J and therefore intensities in the form:

$$I(x) = \lambda(x)B_\infty \tag{3.9}$$

A simple linear regression on these darkest pixels in the RGB space yields B_∞.

Since the backscatter is always blue or green, it is also possible to select the darkest pixels of the red channel only. Both methods yields the same result.
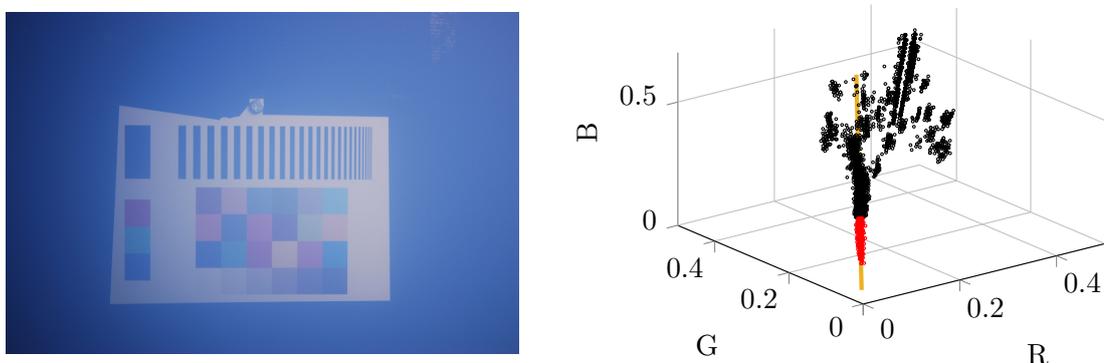
Figure 3.2: (a) Test card image (b) Scatter of all pixels in the RGB space (black), its 10% darkest pixels (red) and the linear regression (orange) yielding B_∞.

3.3.2   Estimation of the backscatter component $B(x)$.

To obtain the scene radiance, we need to suppress the backscatter component $B(x)$ which is a function of the depth and the illumination along the line-of-sight [1]. Since we assumed a non-uniform depth of the scene, $B(x)$ doesn't necessarily vary smoothly so it is estimated by finding the maximum possible value of λ for each pixel knowing that the backscatter is additive. That is, all pixel intensities are orthogonally projected onto the line spanned by B_∞ in the RGB space.

$$B(x) = I(x) \cdot B_\infty \tag{3.10}$$

The scene radiance is then estimated by substracting $B(x)$ from $I(x)$.
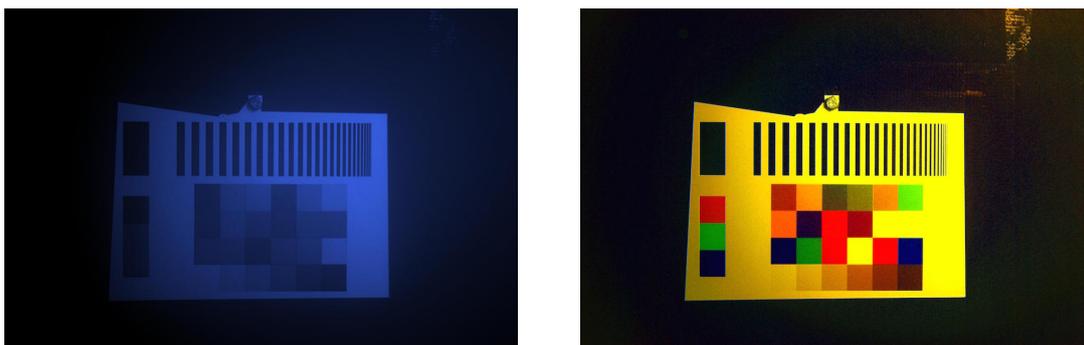
Figure 3.3: (a) Projection of the image onto the line spanned by B_∞. (b) Estimation of the scene radiance.

3.3.3   Performances

Although this method is very simple, it gives acceptable results because it suppress the backscatter in all cases. However, the result is not colour-coherent. Since objects of the scene may contain the haze color, the backscatter is over-estimated. For example the blue color of the haze is also removed from the white parts of the test card which becomes yellow in the radiance estimation of figure 3.3. This method contains only two simple operations: a linear regression over a limited number of dark pixels and an orthogonal projection that makes it suitable for real-time processing.

3.4   Proposed Method for Uniform Depth Images

In this section it is assumed that the images have an approximately uniform depth, often called the planar scene assumption. Although it is a strong constraint, it is a common assumption in the case of ocean floor surveying where the UUV images the scene from above, with relatively small objects and a nearly constant height between the camera and the floor.

3.4.1   Estimation of the backscatter component $B(x)$

B_∞ is estimated as above with a linear regression over the 10% darkest pixels. Now that we assumed a uniform depth scene, the backscatter component $B(x)$ is only a function of the illumination along the line-of-sight. Since illumination vary smoothly as a function of the pixel coordinates, the backscatter also vary smoothly.

We project again the image onto the line spanned by B_∞ in the RGB space and subtract it from $I(x)$. We compute the norm of that quantity to obtain the Euclidean distance map of $I(x)$ from B_∞. Finally that map is thresholded to its 2% smallest values (0.5% to 3% give good results) to obtain the position of all pixels that have the exact color of the backscatter, and therefore are very likely to have a null direct component.

The backscatter is finally recovered by fitting a 2D-polynomial to the previously selected pixels and subtracted to the original image to obtain the radiance.

Similar results can be achieved by guessing the subset of pixels containing only the backscatter over the darkest pixels with a RANSAC algorithm and a cost function that minimize the polynomial regression [1].
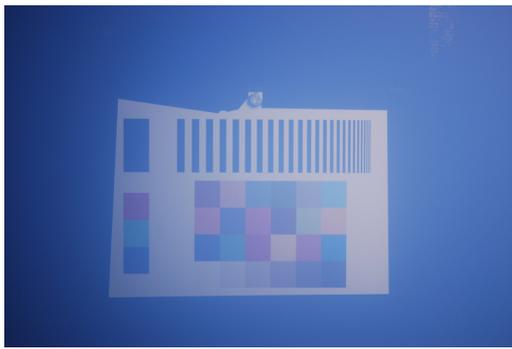
(a) Original image.
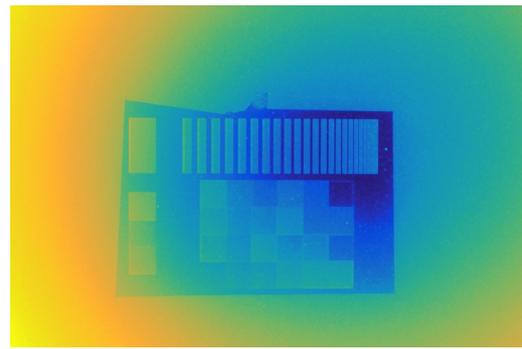
(b) Projection onto the line spanned by B_∞.
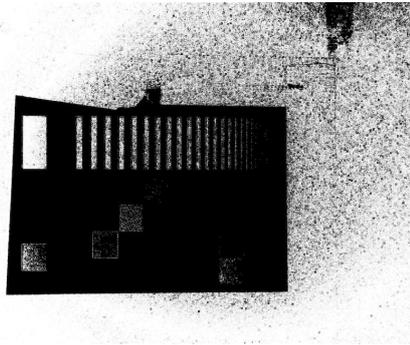
(c) Thresholded distance map to B_∞.
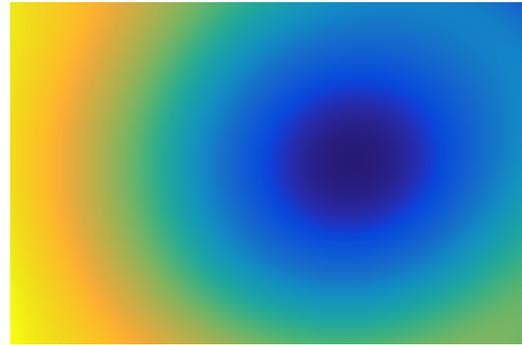
(d) Map of $\lambda(x)$ after 2D-polynomial fitting.
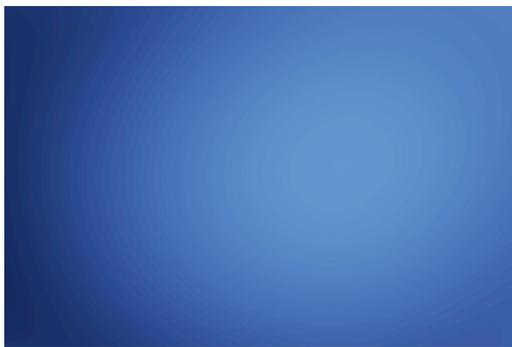
(e) Haze estimation $B(x)$
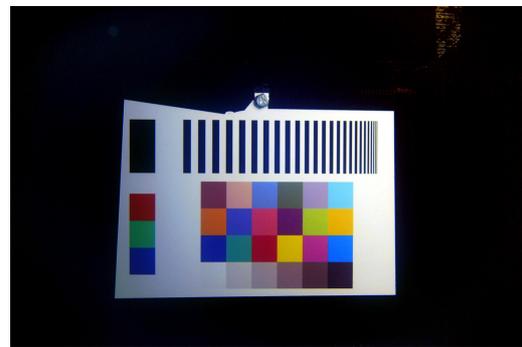
(f) Radiance estimation $J(x)$

Figure 3.4: Steps of the proposed method.

3.4.2   Non-uniform Lighting Compensation

Now that we obtained a good estimation of the scene radiance $J(x)$, we want to tackle the problem of non-uniform lighting. That is, estimate the albedo from the scene radiance. In order to do that, we need a more detailed model of the light behaviour in the image as in [1].

We assume an orthographic camera model with a coordinate system such that the origin is at the camera centre and the xOy plane coincides with the camera sensor. A single artificial point-light source illuminate the environment and is located on the same plane and at a small offset h from the camera.

As always, the intensity measured in each pixel is the sum of the direct component and the backscatter. However, we assume here that the backscatter component of a given pixel is simply the sum of all scattered light within the pixel z-column (line-of-sight) coming directly from the light source, not the light coming from the scene or other pixels column. In other words, we ignore forward scattering and multiple scattering. Forward scattering has been shown to be insignificant, while multiple scattering is known to produce spacial blur [5] and we observe little to none blurring effect in testing images.

Direct Component
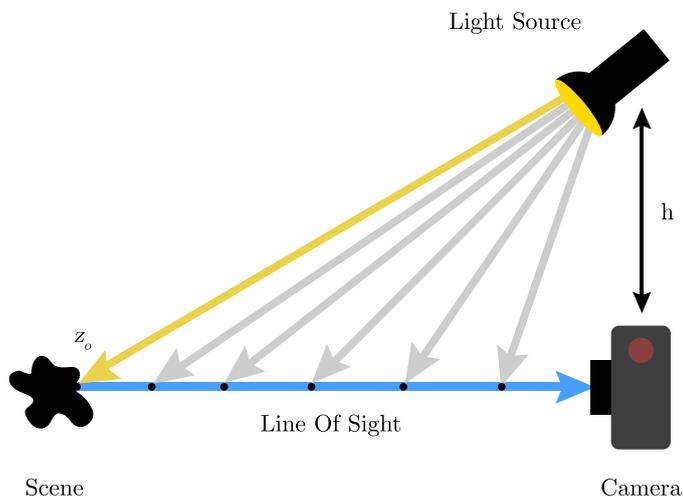
Figure 3.5: Illumination and radiance of the scene object. Adapted from [1].

The Inverse Square Law coupled with the Beer Lambert law gives the illumination intensity of the scene as a function of the distance between the light source and the scene. If I_0 is the source intensity, d the distance between the source and the scene, the illumination of the scene I_S is given by:

$$I_S = \frac{I_0}{d^2} e^{-cd} \tag{3.11}$$

An object situated at distance z_o from the camera is illuminated as in (3.11) with a travelled distance $d = \sqrt{z_o^2 + h^2}$. The object radiance will then be further attenuated by e^{-cz_o} until reaching the camera. Therefore the direct component is given as:

$$D = \frac{I_0}{h^2 + z_o^2} e^{-c(z_o + \sqrt{h^2 + z_o^2})} \rho \tag{3.12}$$

Where ρ is the albedo.

Backscatter Component
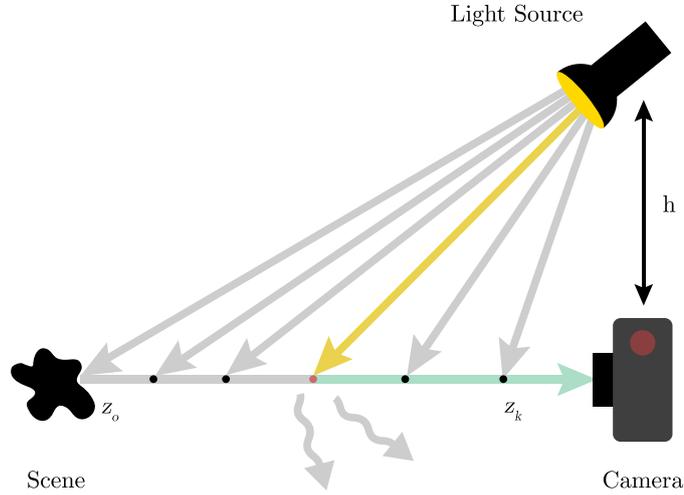
Figure 3.6: Illumination of an elementary volume of medium along the line-of-sight and its backscatter. Adapted from [1].

Considering now a differential volume of depth z withing the (x, y) pixel's column illuminated as before by:

$$I_B = \frac{I_0}{h^2 + z^2} e^{-c\sqrt{h^2 + z^2}} \tag{3.13}$$

A low-order model for the scattering function is given by [8]:

$$\beta(\theta) = \frac{b}{4\pi} \left(1 + g \cos \theta\right) \tag{3.14}$$

The portion $\beta(\theta)$ of I_B that reaches the sensor is again further attenuated by e^{-cz}. Replacing $\cos \theta$ in equation (3.14) by $\frac{z}{\sqrt{h^2 + z^2}}$ yields the intensity contribution of the differential volume:

$$dB(z) = \frac{bI_0}{4\pi} \left(1 + g\frac{z}{\sqrt{h^2 + z^2}}\right) \frac{e^{-c(z + \sqrt{h^2 + z^2})}}{h^2 + z^2} \tag{3.15}$$

As the light carried by the UUV will most certainly be directional, z_k defines the depth from which the cone of light enters the pixel's LOS. We will assume z_k is approximately constant over all pixels. Finally, integrating equation (3.15) from z_k to the scene depth z_0 yields the total backscatter component:

$$B = \int_{z_k}^{z_0} \frac{bI_0}{4\pi} \left(1 + g\frac{z}{\sqrt{h^2 + z^2}}\right) \frac{e^{-c(z+\sqrt{h^2+z^2})}}{h^2 + z^2} \, \mathrm{d}z \tag{3.16}$$

Numerical simulations show that variations of the scattering functions are negligible compared to the attenuation term and that the nearest closed-form solution of B is obtained by approximating the exponential term by e^{-2z}.

$$B \approx \alpha \int_{z_k}^{z_0} \frac{e^{-2cz}}{h^2 + z^2} \, \mathrm{d}z \tag{3.17}$$

With alpha a multiplicative constant. Equation (3.16) and its approximation (3.17) are shown in figure 3.7. The solution of the approximated integral is given by:

$$B \approx \alpha \left[\frac{\mathrm{i}e^{-2\mathrm{i}ch}}{2h} \left(e^{4\mathrm{i}ch} \, \mathrm{Ei}(-2c(\mathrm{i}h + z)) - \mathrm{Ei}(-2\mathrm{i}ch - 2cz) \right) \right]_{z_k}^{z_0} \tag{3.18}$$

Where Ei is the exponential integral function defined as $\mathrm{Ei}(x) = \int_{-x}^{\infty} \frac{e^{-t}}{t} \, \mathrm{d}t$
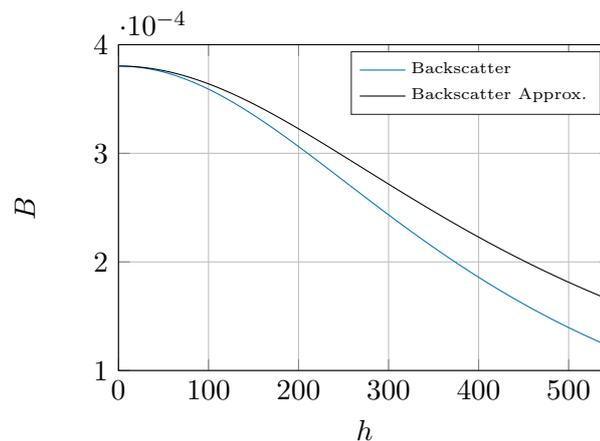
Figure 3.7: Backscatter and its approximation as a function of h.

We are interested in estimating the parameters z_0, z_k and c to compensate the non-uniform lighting but not the multiplicative constant α since color channels can be normalized by their maximum value. Therefore we normalize equation (3.18) by its maximum value. The light being close to the camera, the maximum of $B(x)$ will be within the frame and when that happens, $h = 0$ and equation (3.17) becomes:

$$B \approx \alpha \int_{z_k}^{z_0} \frac{e^{-2cz}}{z^2} \, \mathrm{d}z \tag{3.19}$$

The solution to that integral is given by:

$$B \approx \alpha \left[\frac{2cz \, \mathrm{Ei}(-2cz) + e^{-2cz}}{z} \right]_{z_k}^{z_0} \tag{3.20}$$

The maximum of $B(x)$ is found numerically on the backscatter estimation and we built a distance map $h(x)$ to the maximum. Plotting B as a function of h as in figure 3.8a shows that the orthographic model assumption is not verified in the 4 corners where the pixel's columns are larger and the backscatter augment. Therefore we plot the backscatter as a function of h only on a horizontal line of pixels containing the maximum as in figure 3.8b. Finally the ratio of equations (3.16) and (3.20) is fitted in the least-square sense using Matlab custom equation fitting module to estimate the parameters z_0, z_k and c.
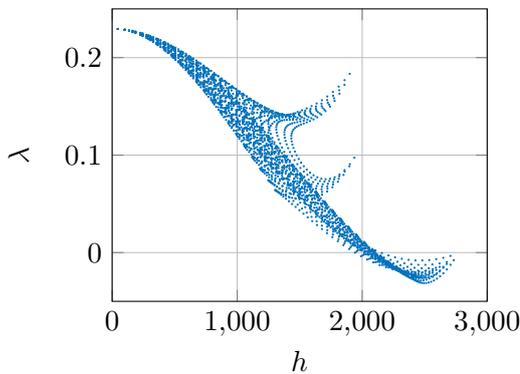
(a) Backscatter as a function of h on the whole image.
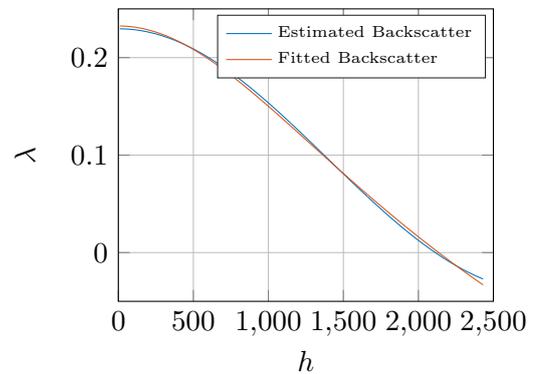
(b) Backscatter as a function of h on an horizontal line of pixel and its fit.

The albedo is finally estimated by multiplying the direct component by the inverse of the attenuation and lighting term and each channel is normalised by its maximum value.

$$I_0\rho = D\frac{h^2 + z_o^2}{e^{-c(z_o+\sqrt{h^2+z_o^2})}} \tag{3.21}$$

(a) Original image.
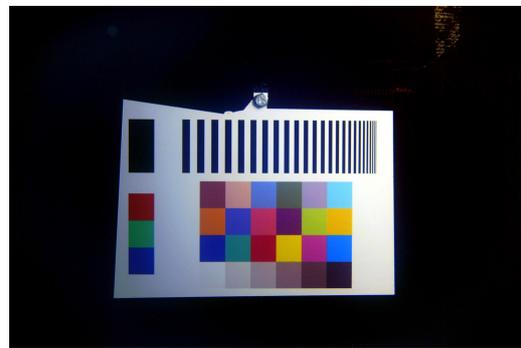
(b) Radiance estimation.
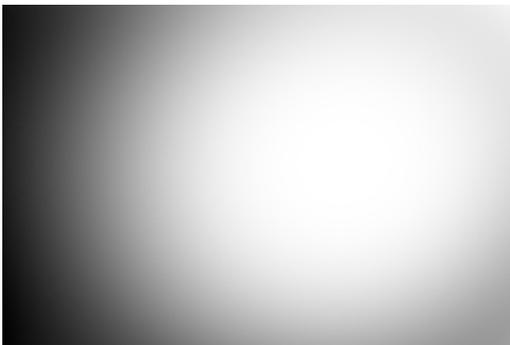
(c) Attenuation and lighting.
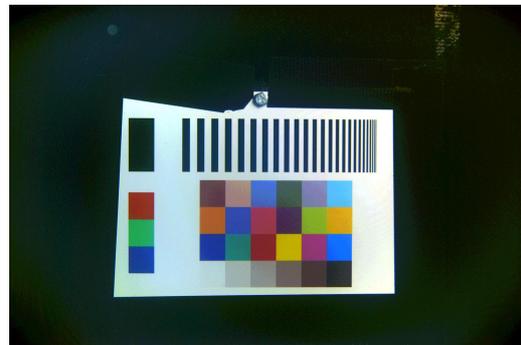
(d) Albedo estimation.

Figure 3.9: Results of the non-uniform compensation on the test card image.

3.4.3 Performances

The method allows for a complete removal of the backscatter component as shown in figure 3.12 that outperforms non-model based methods.

(a) Fusion result. (b) Proposed method albedo.
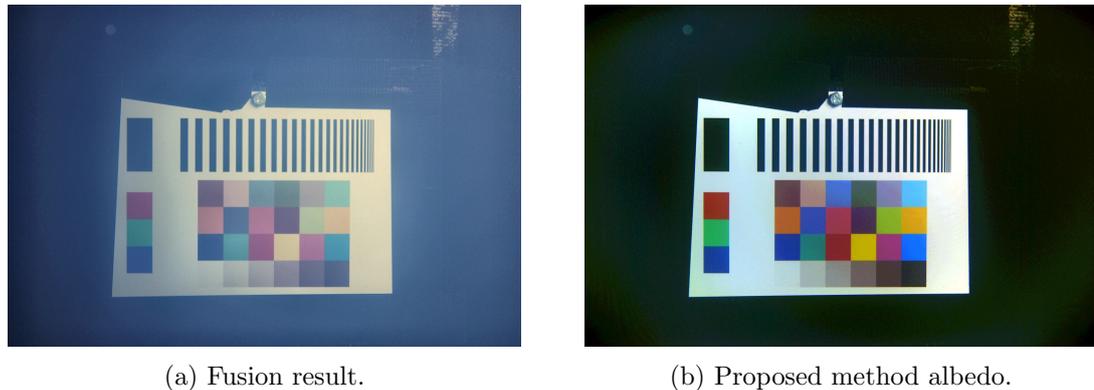
Figure 3.10: Comparison of model and non-model based result.

To the best of our knowledge, there is no existing method that uses the backscatter estimation to compensate for the non-uniform lighting. The resulting albedo estimation yields details that are invisible in the radiance estimation as shown in figure 3.13. Moreover figure 3.11 shows that the albedo estimation of the test card becomes flat and therefore respect the brightness constancy constraint necessary for later mosaicing.

The method uses a linear regression, an orthogonal projection, a 2D polynomial fitting, a least-square custom equation fitting and various additions and multiplications. It has been implemented in Matlab and the processing time is shown in table 3.1. Its computational complexity makes it unsuitable for real-time processing but the processing time could be greatly improved if implemented in OpenCV.

	Radiance Estimation	Albedo Estimation
Test Card (1920x1285)	11.320	0.621
Ocean Floor (1600x1200)	3.158	0.144

Table 3.1: Processing time of the two steps of the proposed method in seconds.

(a) Red Channel Radiance.

(b) Red Channel Albedo.

(c) Green Channel Radiance.

(d) Green Channel Albedo.

(e) Blue Channel Radiance.

(f) Blue Channel Albedo.
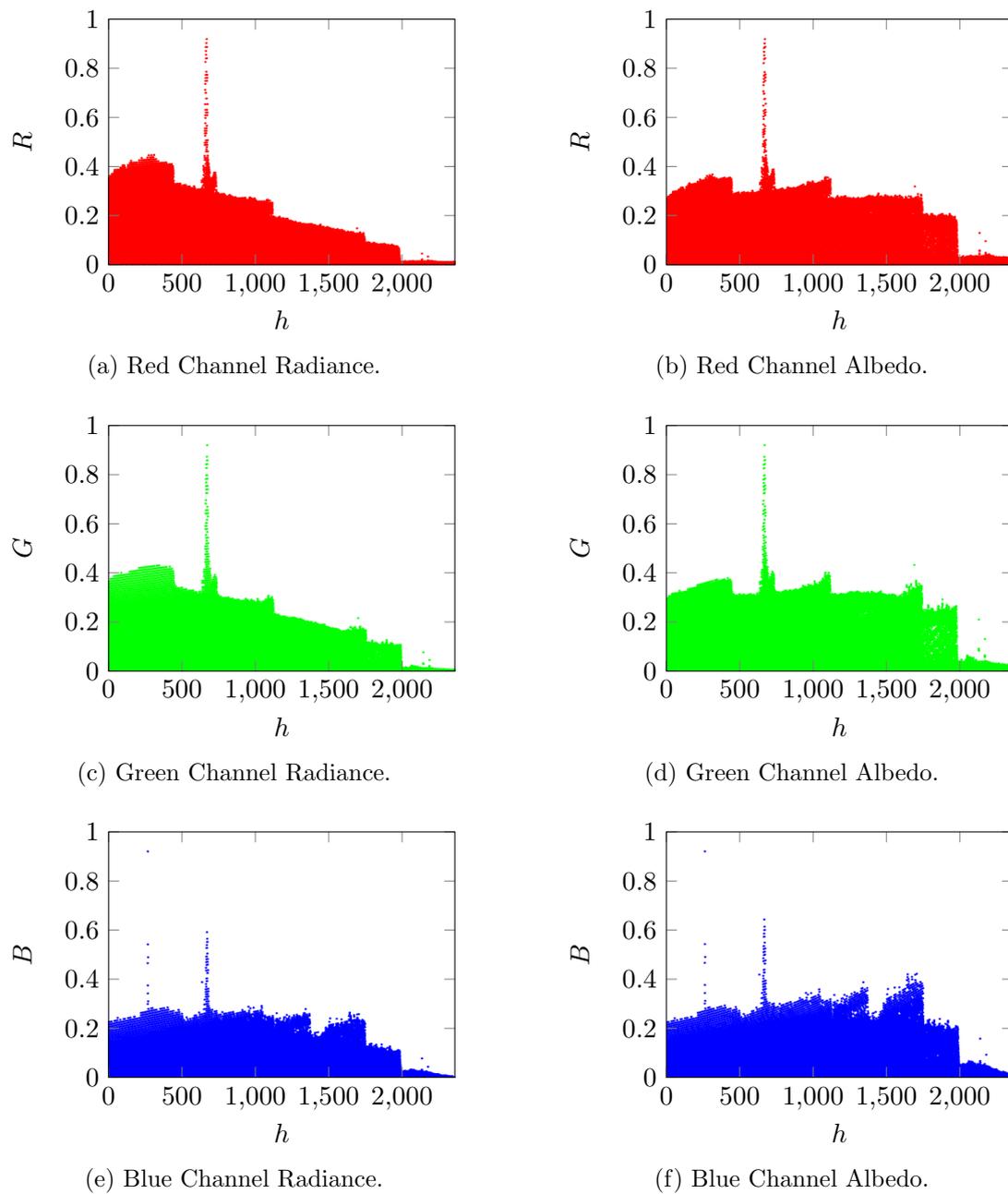
Figure 3.11: Scatters of intensities as functions of h before and after non-uniform lighting compensation.
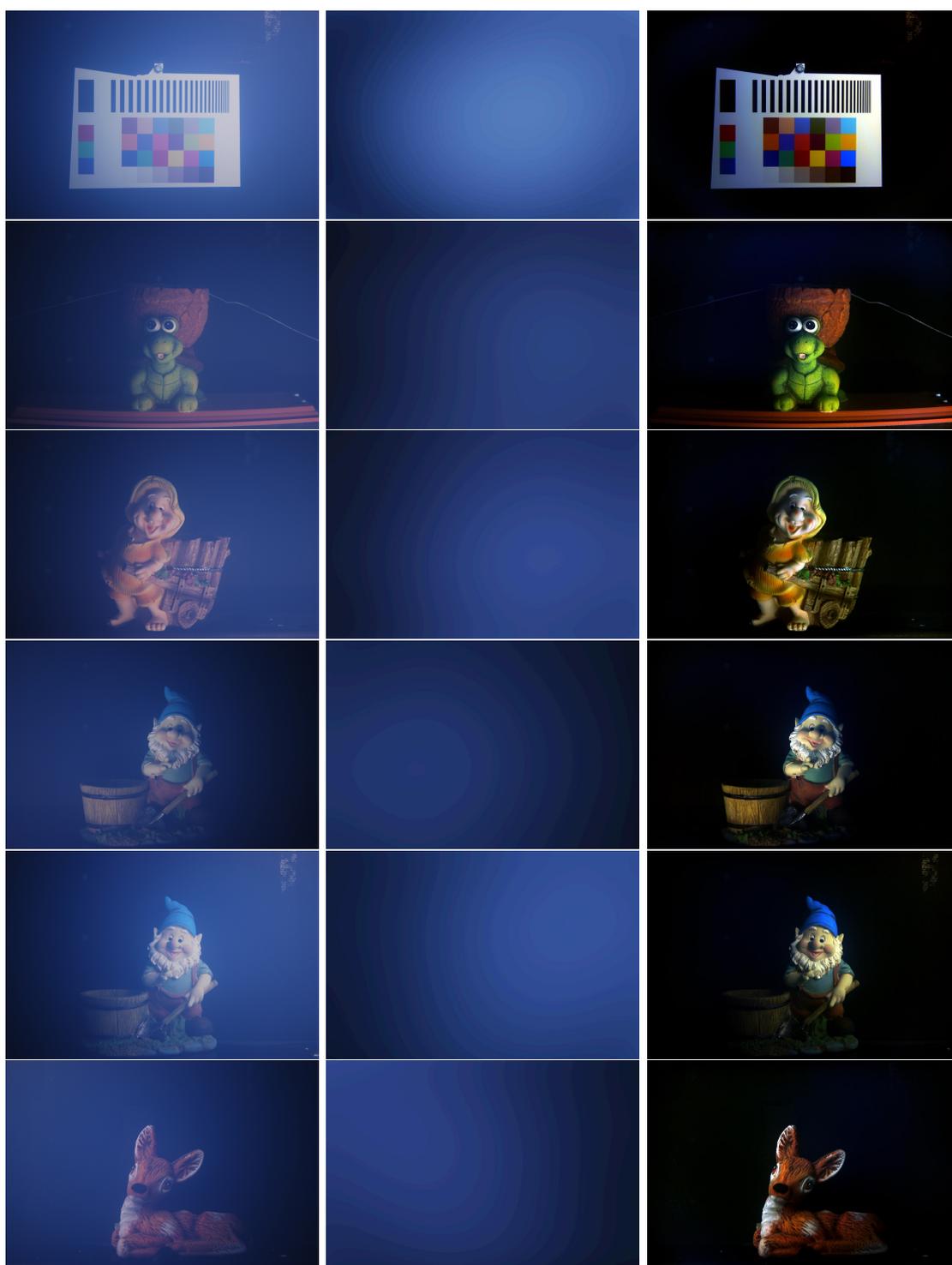
(a) Original. (b) Backscatter. (c) Radiance.

Figure 3.12: Radiance estimation results on a set of testing images.

(a) Radiance.   (b) Albedo.

Figure 3.13: Albedo estimation results on a set of testing images.

3.5   Feature Detection and Matching

Image registration and sparse pose estimation are based on feature detection and matching. Feature extraction typically involves finding well-contrasted patches that are later tracked or matched in another image. In the case of ocean floor surveying where the overlap between successive image is small, matching yields better results that tracking [19]. Once features have been extracted, they can be directly matched using a similarity measure such as the sum of squared differences in intensities. However, a feature description step can be added to make the feature invariant to spatial transformations or illumination changes.

The Scale-Invariant Feature Transform descriptor is widely used for object detection and image registration because of its robustness against scale and orientation changes. The SIFT descriptor is based on the gradient to make the feature insensitive to additive component, and is further improved against illumination changes with a clipping and normalization step [20]. However, our results show that in the case of underwater images the illumination change across images is so strong that a significant part of the matches are lost if applied to the unprocessed images.

False positive matches can be eliminated afterwards in the registration or pose estimation step. Here we show unfiltered matches produced by OpenCV SIFT brute force features matcher. An example on two consecutive images from sequence A is shown in figure 3.14. In all cases, table 3.2 shows the number of matches greatly augment with the radiance estimation, and even further with the albedo estimation of the proposed method. Furthermore, the number of false positives is approximately similar in the radiance and albedo estimations in figure 3.14, but with more total matches for the albedo estimation. Therefore the non-linear illumination compensation effectively diminishes the proportion of false positives.

	Original	Radiance	Albedo
Ocean Floor A	67	310	415
Ocean Floor B	264	433	658
Ocean Floor C	45	87	198

Table 3.2: Number of unfiltered SIFT feature matches.
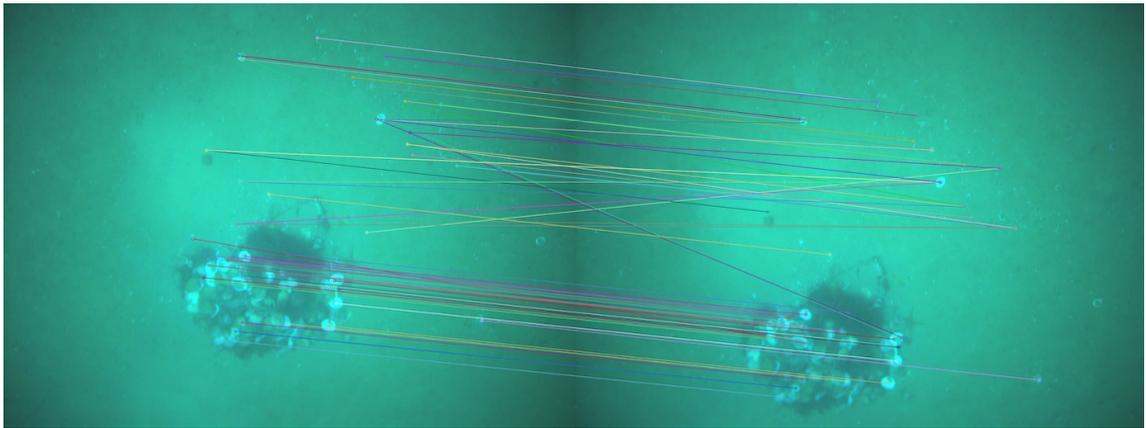
(a) Matching with the original images.
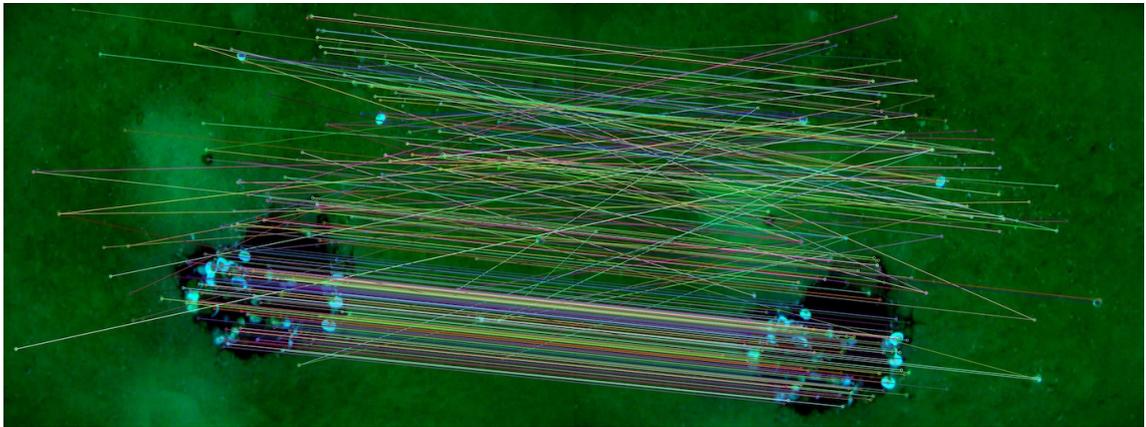
(b) Matching with the radiance estimations.
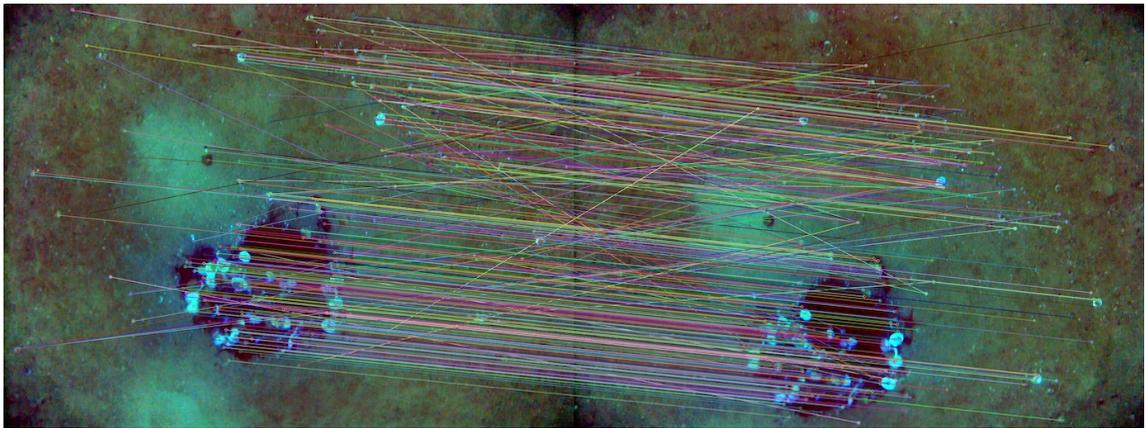
(c) Matching with the albedo estimations.

Figure 3.14: Unfiltered SIFT features matching with 2 consecutive images from ocean floor A.

3.6   Sequence of Images

As mentioned previously the proposed method is particularly adapted for ocean floor surveying with approximately uniform scene depth. For that application, UUVs produce sequence of images that can be mosaicied to gives a global view of the ocean floor. To obtain a good mosaic, images must respect the brightness constancy constraint [6] that is fullfilled by the proposed method after compensating for the non-uniform lighting.

Here we can improve the method by taking into account the added information by the sequence of images. That is, the backscatter is estimated with the darkest pixels of all images instead of one. This yields a more precise estimation of the backscatter constant B_∞, the backscatter regression $B(x)$ and the lighting parameters c, z_0 and z_k.

We apply the method to the 3 testing sets of ocean floor images taken by an AUV and try mosacing with the original images, the estimated radiances and albedos. In every cases the mosaic of albedos estimations gives the most detailed view of the ocean floor. The method shows its limitations with under-exposed or over-exposed corners where the orthographic camera assumption becomes unrealistic. However, the vignetting effect is negligible compared to the results of non-model base methods.

Full size results of the mosaicing are shown in appendix A.

3.6.1   Ocean Floor A
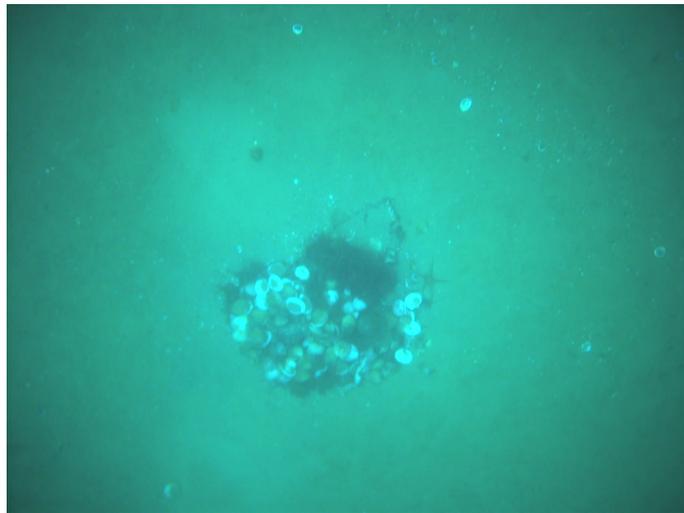
(a) Original image from the sequence.
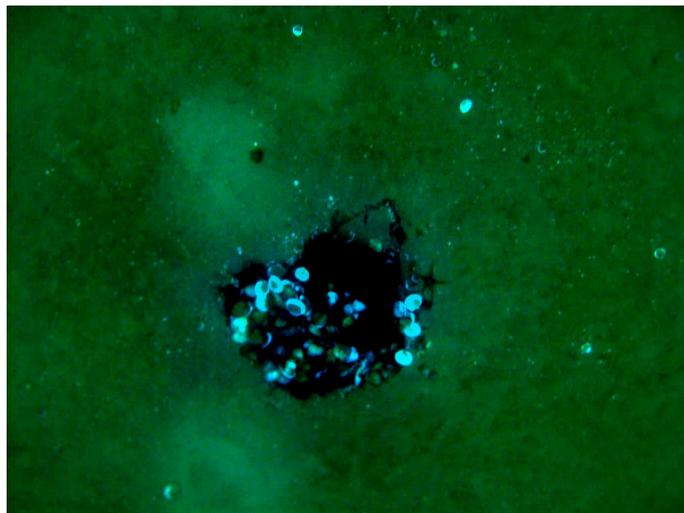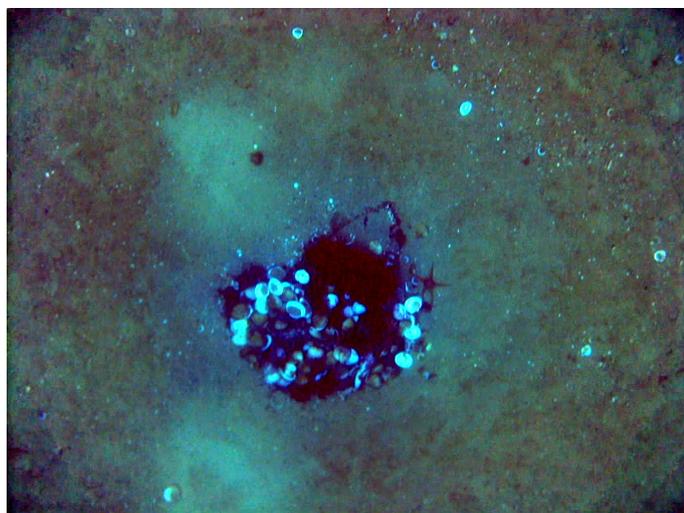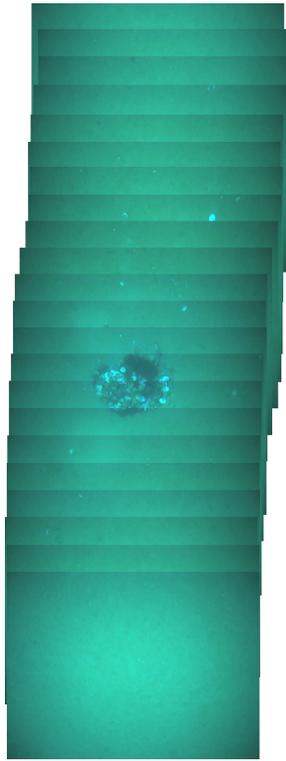
(b) Estimated scene radiance.

(c) Estimated scene albedo.

(a) Direct mosaicing of the sequence.
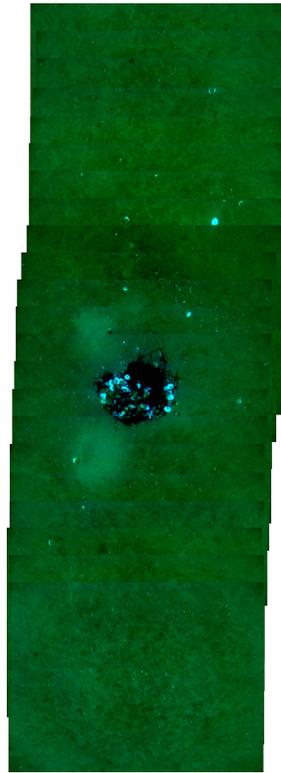
(b) Mosaicing with the estimated scene radiances.
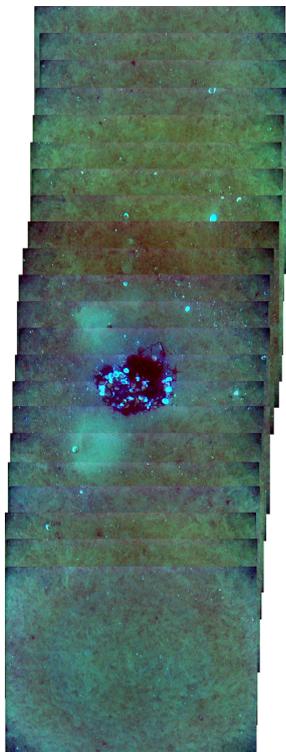
(c) Mosaicing with the estimated scene albedos.
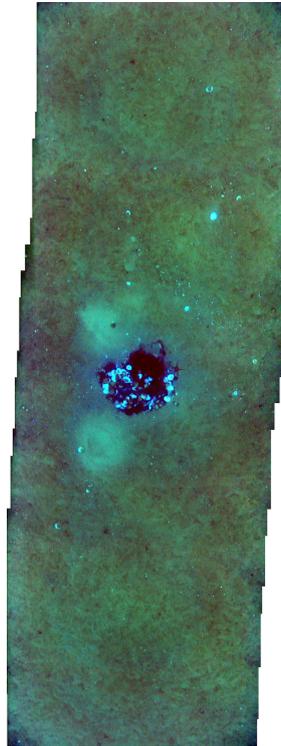
(d) Mosaicing and blending with the estimated scene albedos.

3.6.2 Ocean Floor B
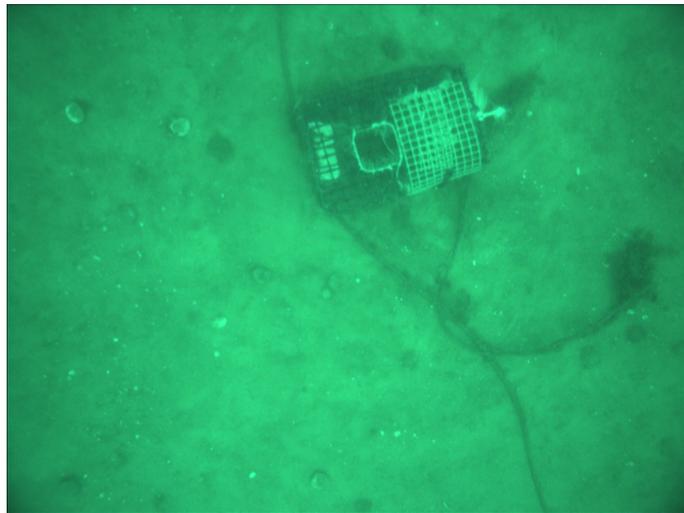
(a) Original image from the sequence.
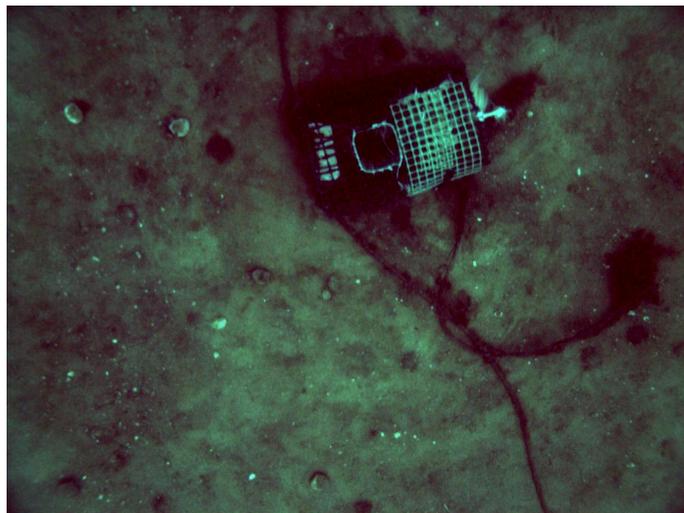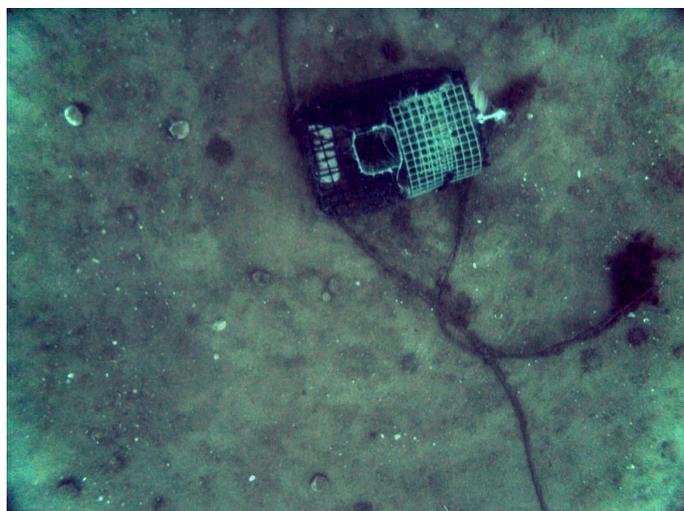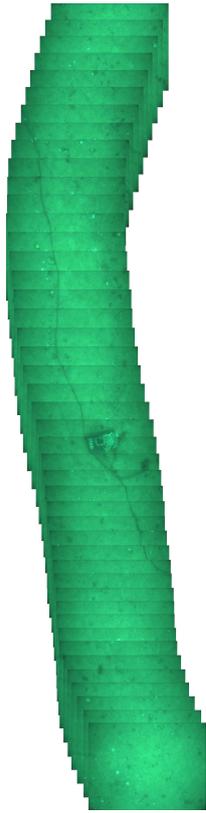
(b) Estimated scene radiance.

(c) Estimated scene albedo.

(a) Direct mosaicing of the sequence.

(b) Mosaicing with the estimated scene radiances.

(c) Mosaicing with the estimated scene albedos.

(d) Mosaicing and blending with the estimated scene albedos.

3.6.3 Ocean Floor C
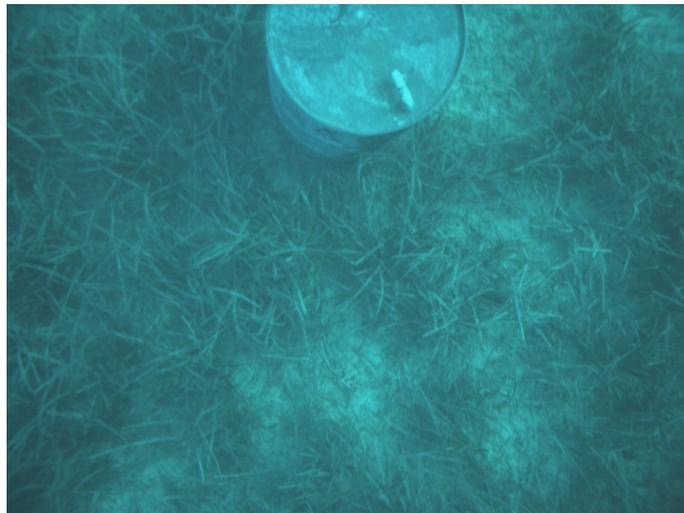
(a) Original image from the sequence.
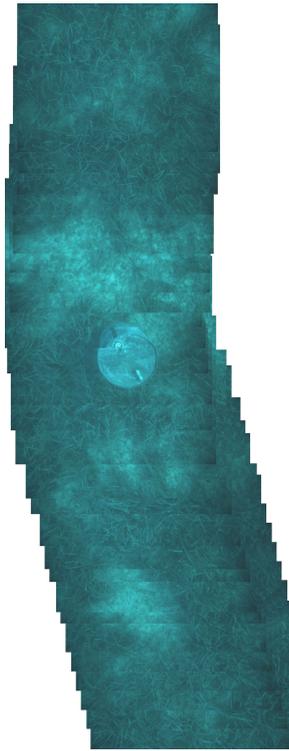
(b) Estimated scene radiance.

(c) Estimated scene albedo.

(a) Direct mosaicing of the sequence.
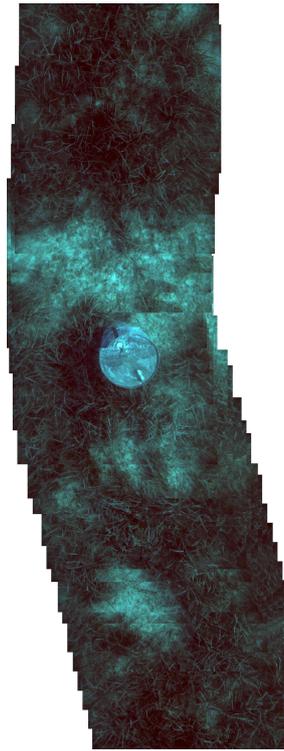
(b) Mosaicing with the estimated scene radiances.
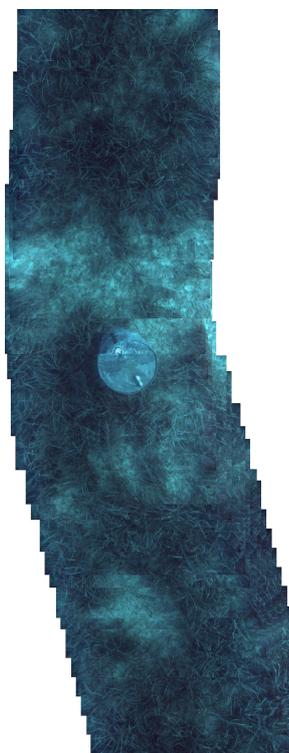
(c) Mosaicing with the estimated scene albedos.
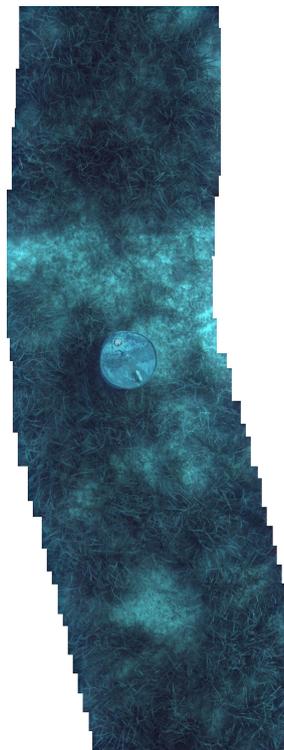
(d) Mosaicing and blending with the estimated scene albedos.

Chapter 4

Conclusion

In this thesis we introduced a new model-based restoration technique specifically designed for artificially illuminated underwater scenes. The results it yields outperform non-model based restoration techniques with a complete removal of the backscatter component. The compensation of the non-uniform lighting allow for better features detection and matching performances, leading to a smooth mosaicing of ocean floor sequences, and possibly enhancing other vision applications such as object detection.

4.1   Future Work

The method has two fixed parameters: the percentage of dark pixels used for the estimation of B_∞ and the threshold value of the distance map from $I(x)$ to B_∞. The two parameters could be automatically estimated, the first one by finding the value that maximise the linearity of the set in the RGB space, and the second by ensuring a uniform spacial distribution of the selected pixels across the image.

The natural next step would be to improve the proposed method for non-uniform depth images. Just as the present method allow for good image registration and mosaicing, an improved method for non-planar scene would allow for proper structure from motion. The method can also greatly be improved in term of processing speed. Real-time estimation of the scene albedo with GPU implementation could be the pre-processing step

allowing for high quality simultaneous localization and mapping.

Finally, temporal coherence could be investigated with an auto-regressive estimation of the backscatter parameters (B_∞ and $B(x)$) and lighting (z_0, z_k and c).

Appendices

Appendix A

Full Size Mosaicing Results
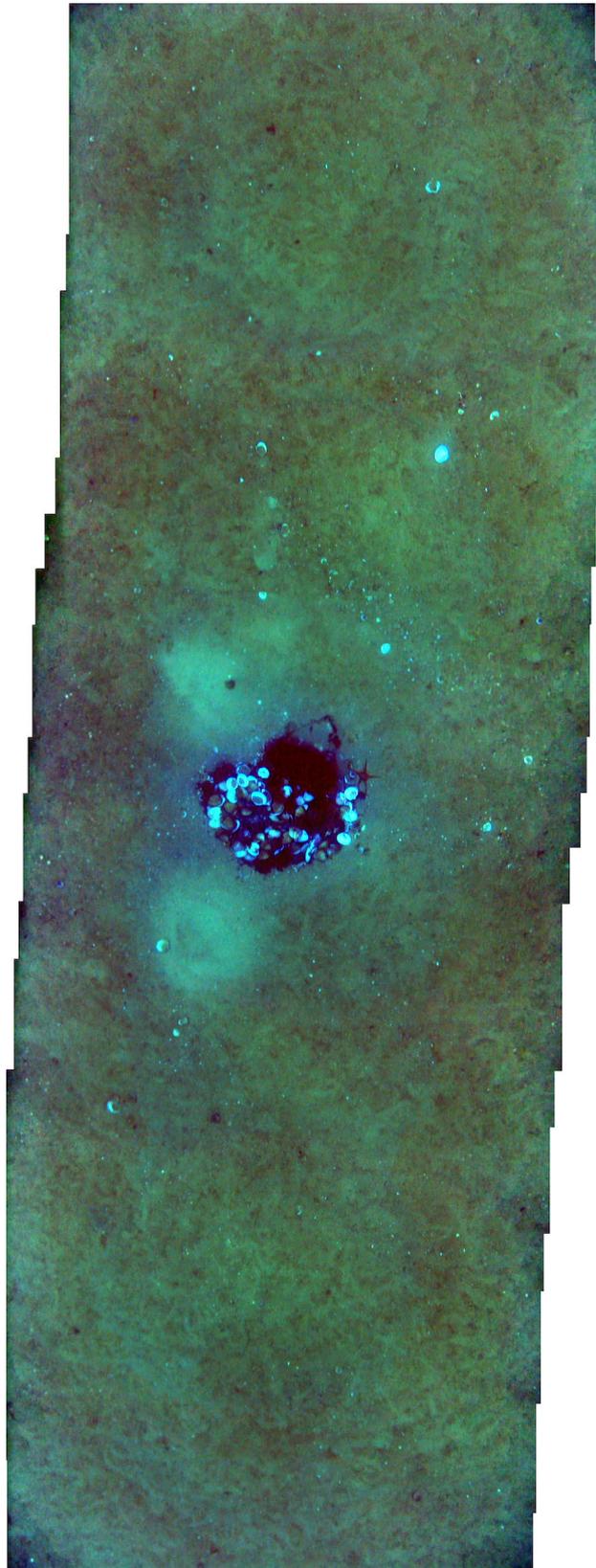
Figure A.1: Mosaicing with the estimated scene albedos of ocean floor A.

Figure A.2: Mosaicing with the estimated scene albedos of ocean floor B.
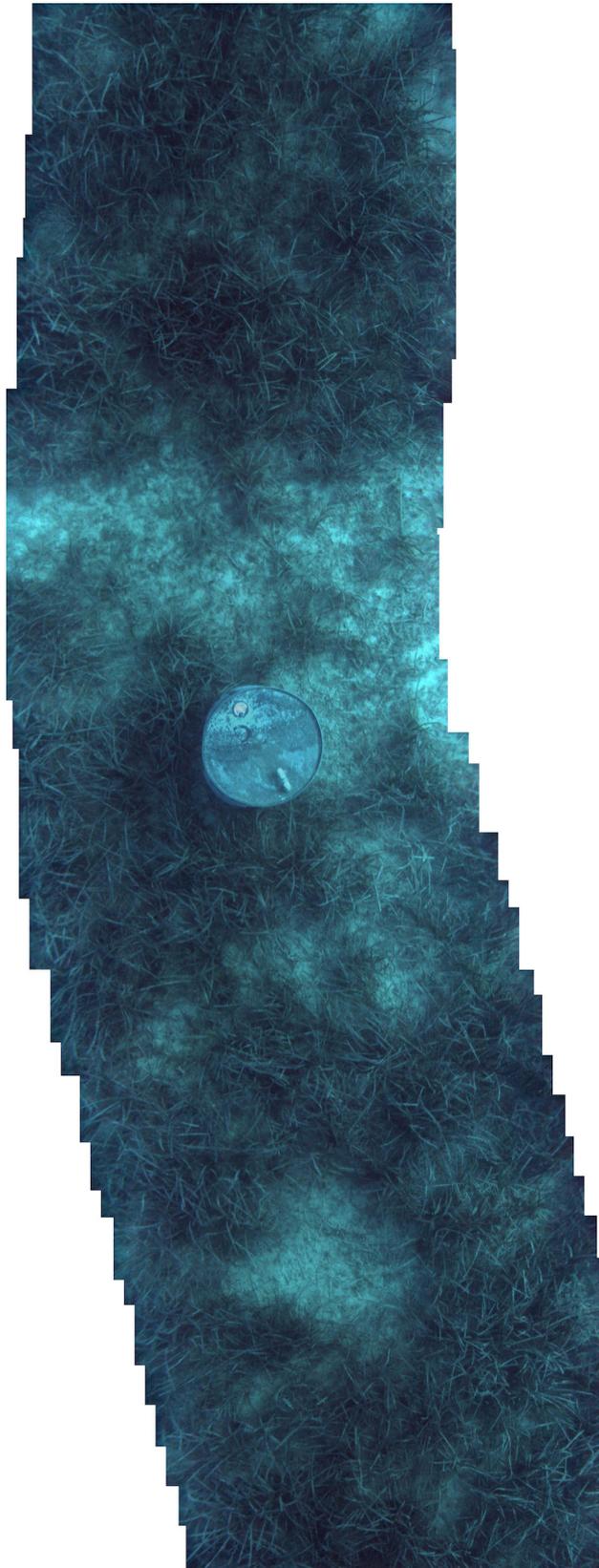
Figure A.3: Mosaicing with the estimated scene albedos of ocean floor C.

Bibliography

[1] C. Tsiotsios, M. E. Angelopoulou, T.-K. Kim, and A. J. Davison, "Backscatter compensated photometric stereo with 3 sources," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on.* IEEE, 2014, pp. 2259–2266.

[2] D. R. Yoerger, A. M. Bradley, B. B. Walden, M.-H. Cormier, and W. B. Ryan, "Fine-scale seafloor survey in rugged deep-ocean terrain with an autonomous robot," in *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, vol. 2. IEEE, 2000, pp. 1787–1792.

[3] N. Gracias and J. Santos-Victor, "Underwater mosaicing and trajectory reconstruction using global alignment," in *OCEANS, 2001. MTS/IEEE Conference and Exhibition*, vol. 4. IEEE, 2001, pp. 2557–2563.

[4] O. Pizarro and H. Singh, "Toward large-area mosaicing for underwater scientific applications," *Oceanic Engineering, IEEE Journal of*, vol. 28, no. 4, pp. 651–672, 2003.

[5] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *International Journal of Computer Vision*, vol. 48, no. 3, pp. 233–254, 2002.

[6] O. Pizarro, "Large scale structure from motion for autonomous underwater vehicle surveys," Ph.D. dissertation, Massachusetts Institute of Technology, 2004.

[7] R. Eustice, O. Pizarro, H. Singh, and J. Howland, "Uwit: Underwater image toolbox for optical image processing and mosaicking in matlab," in *Underwater Technology, 2002. Proceedings of the 2002 International Symposium on.* IEEE, 2002, pp. 141–145.

[8] S. G. Narasimhan and S. K. Nayar, "Structured light methods for underwater imaging: light stripe scanning and photometric stereo," in *OCEANS, 2005. Proceedings of MTS/IEEE.* IEEE, 2005, pp. 2610–2617.

[9] Y. Y. Schechner and N. Karpel, "Clear underwater vision," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 1. IEEE, 2004, pp. I–536.

[10] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 12, pp. 2341–2353, 2011.

[11] N. Cohen, "A color balancing algorithm for cameras," *EE368 Digital Image Processing*, 2011.

[12] G. Buchsbaum, "A spatial processor model for object colour perception," *journal of the Franklin institute*, vol. 310, no. 1, pp. 1–26, 1980.

[13] E. H. Land *et al.*, *The retinex theory of color vision.* Scientific America., 1977.

[14] G. D. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Color and Imaging Conference*, vol. 2004, no. 1. Society for Imaging Science and Technology, 2004, pp. 37–41.

[15] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on.* IEEE, 2012, pp. 81–88.

[16] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics gems IV.* Academic Press Professional, Inc., 1994, pp. 474–485.

[17] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on.* IEEE, 2009, pp. 1597–1604.

[18] K. He, J. Sun, and X. Tang, "Guided image filtering," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 6, pp. 1397–1409, 2013.

[19] R. Szeliski, *Computer vision: algorithms and applications.* Springer Science & Business Media, 2010.

[20] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.