

True Colour Correction of Autonomous Underwater Vehicle Imagery

Mitch Bryson

Australian Centre for Field Robotics
The University of Sydney, NSW, Australia
m.bryson@acfr.usyd.edu.au

Matthew Johnson-Roberson

Department of Naval Architecture and Marine Engineering
University of Michigan, Ann Arbor, MI, USA

Oscar Pizarro

Australian Centre for Field Robotics
The University of Sydney, NSW, Australia

Stefan B. Williams

Australian Centre for Field Robotics
The University of Sydney, NSW, Australia

Abstract

This paper presents an automated approach to recovering the true colour of objects on the seafloor in images collected from multiple perspectives by an Autonomous Underwater Vehicle (AUV) during the construction of 3D seafloor models and image mosaics. When capturing images underwater, the water column induces several effects on light that are typically negligible in air such as colour-dependent attenuation and backscatter. AUVs

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/rob.21638

must typically carry artificial lighting when operating at depths below 20-30m; the lighting pattern generated is usually not spatially consistent. These effects cause problems for human interpretation of images, limit the ability of using colour to identify benthic biota or quantify changes over multiple dives and also confound computer-based techniques for clustering and classification. Our approach exploits the 3D structure of the scene generated using structure-from-motion and photogrammetry techniques to provide basic spatial data to an underwater image formation model. Parameters that are dependent on the properties of the water column are estimated from the image data itself, rather than using fixed in-situ infrastructure such as reflectance panels or detailed data on water constituents. The model accounts for distance-based attenuation and backscatter, camera vignetting and the artificial lighting pattern, recovering measurements of the true colour (reflectance) and thus allows us to approximate the appearance of the scene as if imaged in air and illuminated from above. Our method is validated against known colour targets using imagery collected in different underwater environments by two AUVs that are routinely used as part of a benthic habitat monitoring program.

1 Introduction

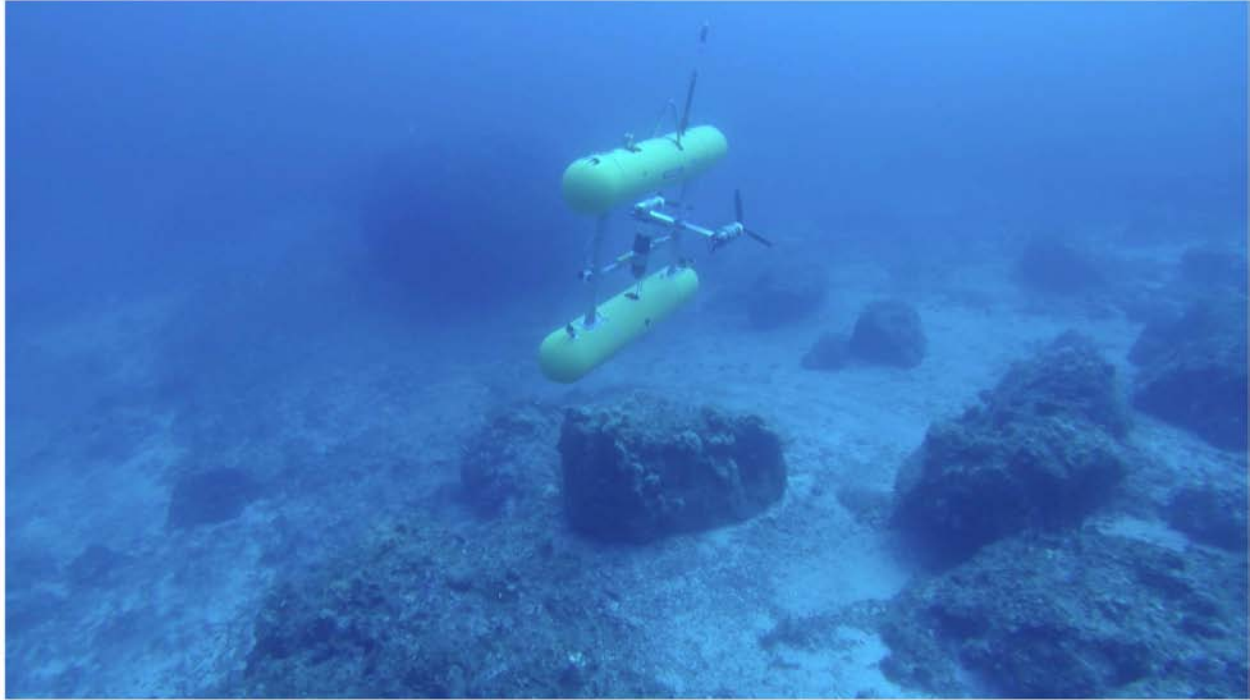
In recent years, marine biologists and ecologists have increasingly relied on imagery from platforms such as Autonomous Underwater Vehicles (AUVs) for monitoring marine benthic habitats such as coral reefs, kelp forests and seagrass meadows (Clarke et al., 2009; Grasmueck et al., 2006; Smale et al., 2012; Yoerger et al., 2007). Underwater imagery can be used to classify and count the abundance of various species in an area and data collected over multiple sampling times can be used to infer changes to the environment and population, for example due to pollution, bio-invasion or climate change (Bewley et al., 2012; Bryson et al., 2013).

When capturing images underwater, the water column induces several effects on images that are not typically seen when imaging in air. Water causes significant attenuation of light passing through it, reducing its

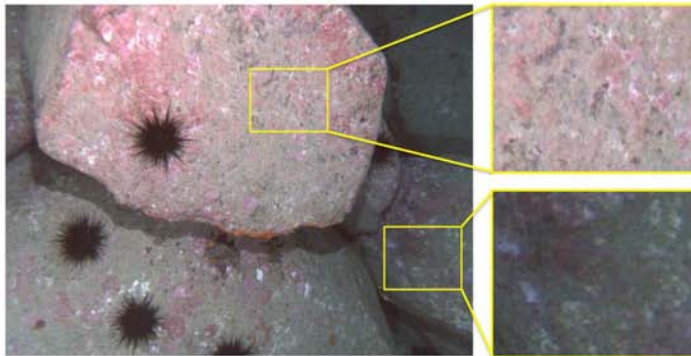
intensity exponentially with the distance travelled (Jaffe, 1990). For this reason, sunlight, commonly used as a lighting source in terrestrial photogrammetry, is typically not strong enough to sufficiently illuminate scenes below depths of approximately 20-30m, necessitating the use of artificial lighting on-board an imaging platform. The attenuation of light underwater is frequency-dependent; red light is attenuated over much shorter distances than blue light, resulting in a change in the observed colour of an object at different distances from the camera and light source (for example see Figure 1). Additionally, the magnitude of attenuation and scattering depend on several complex factors including water temperature and salinity, water chemistry and the type and quantity of particulates in the water such as plankton (Jaffe, 1990; Mobley, 1994). The resulting colour and brightness of objects measured in images varies significantly when imaged from different camera perspectives and distances from a moving platform such as an AUV. This imagery is typically difficult to interpret by end-users and causes issues for quantitative computer-based techniques for clustering and classification based on colour (Beijbom et al., 2012; Bewley et al., 2012; Soriano et al., 2001).

The aim of this paper is to develop a methodology for recovering the true colour of objects on the seafloor captured in underwater images (in contrast to previous work on improving colour balance or consistency (Bryson et al., 2012; Torres-Mendez and Dudek, 2005; Vasilescu et al., 2010)) by removing the perspective dependent effects induced by attenuation and scattering in the water column and the fixed lighting pattern carried by the AUV. By “true colour”, we refer to the reflectance or albedo of points on the seafloor corresponding to the three channels of the colour imaging camera. Using no information about the spectral response function of the camera or spectral colour of the illuminating light source, our method allows for the recovery of images that would be gained by imaging the surface in air (where attenuation and scattering are negligible over short distances) and where the lighting is from a fixed direction with respect to the seafloor surface (i.e. from above). When knowledge of the spectral response of the camera and spectral colour of the lights is available, our method additionally allows for the recovery of reflectance or albedo measurements of the surface such that images can be constructed as if in air and illuminated by a white light source, allowing for the quantitative comparison of surfaces imaged at different times by different cameras and light sources. The resulting colour of objects in both cases is no longer dependent on the perspective from which a particular element of the surface was imaged from, allowing for quantitative comparison of images taken from

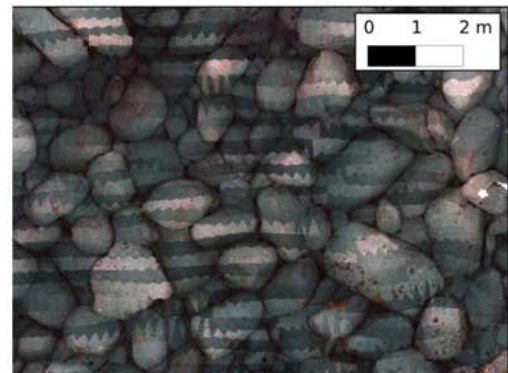
different perspectives and thus potentially over multiple AUV deployments. Additionally, measurements of accurate colour are useful as a feature for distinguishing between coral species (Hochberg and Atkinson, 2000) and other benthic organisms and have been shown to be related to other biophysical properties such as chlorophyll pigment composition and benthic organism health (Hochberg et al., 2006).



(a) Imaging over rugous 3D underwater terrain from an AUV



(b) Example downwards looking image from the AUV with highlighted sections of rocky surface in different portions of the image.



(c) Section of a mosaic constructed using multiple images collected by the AUV. The horizontal banding of light and dark corresponds to variations in the altitude of the AUV during multiple passes over the terrain.

Figure 1: Underwater Imaging using an Autonomous Underwater Vehicle (AUV): Vehicle-fixed lighting patterns and water-based light attenuation cause foreground objects to appear bright and red with respect to background objects that appear dark and blue/green. Inconsistencies in the colour and brightness in images are visible across a single image (b) and when multiple images are combined into a single mosaic (c) these result in visual artifacts.

Our method exploits known structure of the 3D landscape and the relative position of the camera, derived using structure-from-motion and photogrammetry (Johnson-Roberson et al., 2010; Mahon et al., 2008), to provide the necessary inputs to an underwater image formation model including lighting pattern, frequency-dependent attenuation and backscatter. Our approach estimates water column attenuation and backscatter parameters from the images themselves (in contrast to previous work which estimates these values from in-situ calibration infrastructure (Sedlazeck et al., 2009; Yamashita et al., 2007)). When imaging in large-scale biological habitats, the use of colour-calibration tools such as colour-boards is undesirable due to logistical complexity and the sensitivities of marine habitats to man-made disturbances to the seafloor. Additionally our method allows for the processing and correction of historical imagery that lacks in-situ colour calibration data. Results of the method are illustrated in two different AUV dives over different benthic habitat types and ocean conditions.

Section 2 provides a review of underwater imaging and past work in underwater image colour correction. Section 3 presents our methodology including a review of underwater 3D reconstruction, our image formation model and techniques used to estimate image formation model parameters and how these are used to recover true colour. An overview of our experimental setup is provided in Section 4 with results presented in Section 5. Discussion, conclusions and future work are presented in Sections 6 and 7.

2 Background Literature

2.1 Underwater Image Formation

The measured image intensity recorded by a camera at a point in an underwater scene is not directly proportional to the actual ‘brightness’ and ‘colour’ of the point; instead, several factors act on the light path that make this relationship dependent on the water properties and the perspective of the camera and lights with respect to the point. The predominant effects of the water column on light used for underwater imaging are scattering and attenuation (Jaffe, 1990; Mobley, 1994). Attenuation occurs when light is absorbed or diffracted by water molecules or other particles in the water (Jaffe, 1990) and is dependent on water

temperature, salinity, water quality (i.e. from pollution, sediment suspension) and suspended microscopic life (plankton). Attenuation is the dominant cause of the colour imbalance often visible in underwater images.

In addition to attenuation, light is scattered during its travel through the water, i.e. some proportion of the light power is redirected away from the principle direction of travel, owing to interactions with small, particulate matter in the water column. Scattering is typically split into two components, forward scatter and backscatter (Jaffe, 1990; Lewis et al., 1999), which occur due to identical physical processes but have differing implications in regards to underwater imaging. Forward scattering occurs when light heading towards the direction of the camera is refracted by very small angles through the water, resulting in blurring and a lack of contrast in collected images. Backscattering occurs when light not heading towards the camera is reflected (typically by larger angles) back towards the camera's line of sight, resulting in a visible bright haze in images.

2.2 Underwater Image Processing

There is a large body of existing work on correcting images for the effects of the underwater environment; for an extensive review, the reader is referred to (Schettini and Corchs, 2010). The authors in (Chiang and Chen, 2012) proposed a method for de-hazing or removing the effect of colour change in underwater images using a combination of filtering and approximation of the scene depth from the intensity of image data. Related approaches to the removal of haze from terrestrial images (for example from fog or pollution (Tarel and Hautiere, 2009)) have also been used in the correction of underwater images (Gracias et al., 2013).

The authors in (Schechner and Karpel, 2004) present an approach to reducing the effects of light scattering by taking several images through a polarising filter at various angles. The method requires several images of the same scene measured at different polarising filter orientations and is suited to imaging from a static platform and in a horizontal direction.

Several authors have proposed approaches to compensating for the effect of attenuation on image colour. In (Vasilescu et al., 2010), the authors present an active imaging strategy that adaptively illuminates a scene

during imaging based on the average depth from the camera. The approach alters the colour balance of the strobe lighting source (for example to increase red-light for scenes further from the camera) but uses only one depth value (derived from the camera focal length) per scene, neglecting the case where different objects in a single scene are observed at different ranges to the camera. In (Yamashita et al., 2007), the authors propose a method for correcting underwater images for colour-attenuation by estimating attenuation coefficients using either a single image captured both underwater and in air or two underwater images captured at different depth values. The method relies on a controlled setup for capturing images and is only demonstrated on three images. Similarly, the authors of (Sedlazeck et al., 2009) present a method for colour correcting underwater images for attenuation using known terrain structure under the assumption that a sufficient number of ‘white’ points can be identified in the data, either from known properties of the scene or by using calibration targets. While these approaches are insightful, they are typically impractical in large-scale, unstructured underwater environments. In (Torres-Mendez and Dudek, 2005), the authors present a learning-based approach to colour correction that learns the average relationship between colours underwater and in air using training points of images of the same objects both in and out of water. The approach ignores the effect of distance-based attenuation, instead learning a single transformation that corresponds to an object at a fixed imaging distance. These past approaches to underwater colour correction either ignore the explicit causes of attenuation or require complicated and limiting calibration setups for attenuation-compensation which are logistically impractical in large-scale surveying in unstructured underwater environments.

2.3 Underwater Imaging using AUVs

AUVs are increasingly being used for oceanic and seafloor research and monitoring as a replacement for traditional means of exploration such as SCUBA diving and ship-towed video. AUVs are particularly well suited to tasks such as multibeam sonar surveying and high-resolution optical imaging of the seafloor due their ability for precise trajectory control and reduced reliance on a surface vessel (when compared to Remotely Operated Vehicles (ROVs)). AUVs provide the ability to collect large numbers of potentially overlapping images that are time-registered to other navigation sensor data collected by the AUV. The authors in (Kaeli et al., 2011), for example, use data from a doppler velocity log to approximate the planar depth of a camera

mounted to a towed robotic system that was then used to correct for the lighting beam pattern in underwater images. The sensor data collected by AUVs allows for the recovery of 3D and other spatial information using techniques in structure-from-motion, photogrammetry and Simultaneous Localisation And Mapping (SLAM) (Campos et al., 2011; Campos et al., 2015; Eustice et al., 2008; Johnson-Roberson et al., 2010; Mahon et al., 2008). For example, the Australian Centre for Field Robotics operates two ocean-going AUVs, *Iver* and *Sirius* (see Figure 5) that are capable of undertaking high-resolution, geo-referenced surveys and are currently used as part of Australia’s Integrated Marine Observing System (IMOS) (Williams et al., 2012). As part of the IMOS program, *Sirius* is deployed at several key locations along Australian coastal waters on a yearly basis to perform repeated surveys and collect data which can be used for long-term monitoring. Both platforms use stereo image data and other navigation sensor information onboard the AUV to build three-dimensional (3D) reconstructions of the underwater terrain associated with the imagery (Johnson-Roberson et al., 2013; Johnson-Roberson et al., 2010; Mahon et al., 2008).

The ability to produce this information co-registered to the collected imagery provides most of the input required for explicitly correcting for spatially-dependent underwater image formation effects (such as attenuation). The authors in (Sedlazeck et al., 2009) utilise these techniques to build 3D models that provide data to an inverse image formation model, but still require additional assumptions about the colours of objects in the environment (i.e. white points). In the authors’ previous work (Bryson et al., 2012), we developed a method for improving colour consistency in multi-perspective underwater imaging by using a colour balancing algorithm that accounted for attenuation. This algorithm balanced the gain between the red, green and blue channels of the image (using structure-from-motion-derived scene ranges) by assuming that the overall distribution of colours in the world was gray and that colours were distributed evenly over different ranges to the camera. While providing colour-consistency, this algorithm did not provide true colour per-se and exhibited poor performance in environments in which the spatially-even grayworld assumption was violated.

In contrast to past work, the contribution our work in this paper is the development of a method for recovering true colour from underwater images by explicitly accounting for spatially-dependent underwater image formation without requiring an in-situ colour calibration setup.

3 Methodology

This section of the paper provides a background in underwater structure-from-motion, presents our underwater image formation model and describes our approach to underwater image correction.

3.1 Underwater Structure from Motion

The colour correction strategy discussed throughout the rest of the paper exploits knowledge about precise distances between the camera and each point in an observed scene. This information can be gained from a variety of different methods that use either stereo images or sufficiently overlapping monocular images (see Section 2.3 above). In this paper, this information is generated using an existing structure-from-motion processing pipeline (Johnson-Roberson et al., 2013; Johnson-Roberson et al., 2010; Mahon et al., 2008). The processing pipeline is made up of the following steps:

1. **Data collection:** Overlapping stereo-image pairs are collected over an underwater scene using an AUV. Additional navigation sensor data is also collected (a depth sensor, doppler-velocity sensor and attitude sensors) to assist in localisation and mapping accuracy over long-distance transects made by the AUV.
2. **Image Feature Processing:** Scale Invariant Feature Transform (SIFT) feature points (Lowe, 2004) are extracted from each stereo image pair and used to compute pose-relative 3D point features. These pose-relative point features are also matched and tracked across overlapping poses to generate pose-to-pose relative position and orientation constraints.
3. **Pose Recovery:** The pose-to-pose relative constraints are used in a Simultaneous Localisation and Mapping (SLAM) algorithm (Mahon et al., 2008) with additional navigation sensor information to compute the trajectory of the platform and precision three-axis orientation at each pose.
4. **Terrain Surface Map Recovery:** The estimated trajectory and pose-relative 3D point features are then used to construct a global feature map of the terrain. The point feature map is re-sampled

and triangulated to produce a terrain surface model. Using the estimated poses and 3D terrain surface model it is therefore possible to compute the relative position (and therefore range) of any point on the terrain to the position of the camera and lights during the capture of any of the images in the dataset.

5. **Mosaic Reconstruction and Visualisation:** For visualisation, the surface model is photo-textured by projecting colour camera images onto the estimated terrain structure and orthographic projections of this model are rendered to produce imagery mosaics (Johnson-Roberson et al., 2013; Johnson-Roberson et al., 2010).

Further details on the approach used can be found in (Johnson-Roberson et al., 2013; Johnson-Roberson et al., 2010; Mahon et al., 2008).

3.2 Underwater Image Formation Model

In this section we develop an underwater image formation model that encapsulates the effects of ocean water on light and describes the measured intensity of a point in an image as a function of the albedo (reflectance) of the point. Figure 2 illustrates the overall image formation model which is described in detail in the subsections below. Figure 2 (a) shows the relative position of the AUV, the onboard camera, the onboard lights used to illuminate the imaged terrain and the position of a point of interest seen by the camera. The relative position of the lights, camera and imaged point are all described using vectors that are referenced with respect to an orthogonal coordinate system centred at the focal point of the camera (referred to as the ‘camera frame’, for which vectors are labelled with a superscript c).

3.2.1 Lighting Parameters

Our model considers a number of directional light sources that are attached to the AUV and used to illuminate the imaged scene. We consider the scenario in which the AUV operates at a sufficient depth that these are the only sources of light in the scene (i.e. sunlight does not provide a significant fraction of the illumination

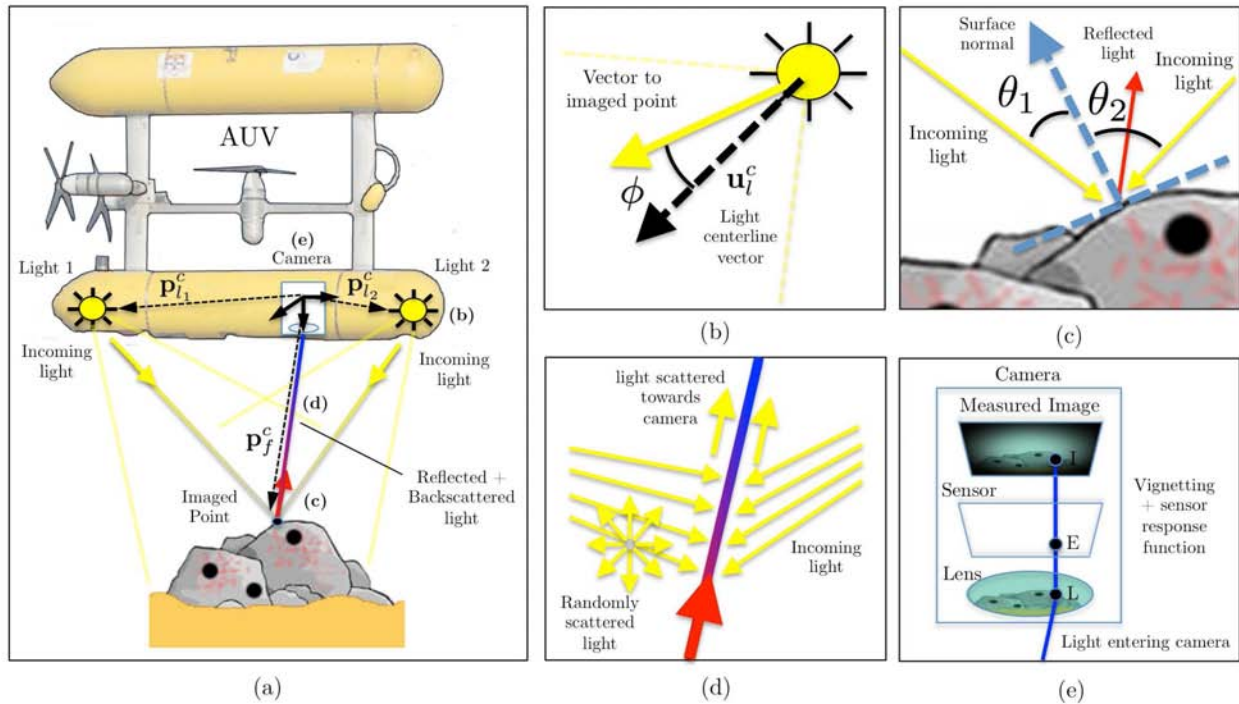


Figure 2: Underwater Image Formation Model: Our model accounts for images of an underwater scene taken from a moving platform such as an AUV (a). Our model accounts for illumination of the scene from a number of light sources, each with a non-uniform spatial pattern (b). Light is attenuated through the water column on its way to the target surface and reflected based on a Lambertian surface assumption (c). Light is further attenuated on its way back to the camera, and the contribution of backscatter in the water between the camera and the target is included (d). Lastly, our model considers the vignetting and sensor response function of the camera (e).

of the surface) or that the platform operates at night. This restriction reduces the complexity of our image formation model and allows the model to be used in situations in which there does not exist accurate and detailed information about external illumination conditions (such as water surface conditions, cloud cover and atmospheric conditions/haze) at the time of imaging. More discussion of this restriction is provided in Section 6 at the end of the paper.

The position of each light with respect to the camera is denoted by the vectors $\mathbf{p}_{i_l}^c$ where $i = 1$ to N_l where N_l is the number of lights. Each light has a centreline vector $\mathbf{u}_{i_l}^c$ which describes the principle pointing direction of the light (see Figure 2 (b)). The power of light emanating from the source is maximal along this vector and drops off in intensity when emanating from directions with an increasing angle ϕ from the centreline. Assuming the use of a Gaussian diffuser in front of the light source (common in underwater applications), the directional reduction in light power is modelled using a Gaussian function:

$$P_\phi(\lambda) = P_0(\lambda)e^{-\frac{1}{2}\frac{\phi^2}{\sigma_l^2}} \quad (1)$$

$$\sigma_l = \sqrt{\frac{\phi_{50\%}^2}{-2\log 0.5}} \quad (2)$$

where $P_0(\lambda)$ is the maximum light power at the centerline, $P_\phi(\lambda)$ is the light power at angle ϕ (both a function of λ , the wavelength of light considered) and $\phi_{50\%}$ is the angle at which the light power drops to 50% of the maximum value. In our experimental setup, $\phi_{50\%} = 40^\circ$, which was determined using information provided by the manufacturer of the light source.

3.2.2 Object Reflection Model

Albedo (sometimes referred to as reflectance) is the ratio of light reflected from a surface to the light falling upon it and is an inherent property of the surface type. For most surfaces, albedo is a function of the wavelength of light considered. In this paper we consider the albedo of an object at the principle wavelengths measured by the camera (i.e. red, green and blue) the combination of which defines the colour and brightness of an object, normalised against the brightness and colour of the illuminating light source.

For general surfaces, the albedo of a surface is dependent on the direction from which light hits the surface and the direction from which the surface is viewed, and is characterised by a Bidirectional Reflectance Distribution Function (BRDF). The use of the BRDF is beyond the scope of our work; instead we only consider Lambertian surfaces (otherwise known as diffuse reflective surfaces) where light is scattered from the surface reflection equally in all directions, and where the albedo is independent of the perspective of the viewing direction and only on the perspective of the illumination source (Pedrotti and Pedrotti, 1993). In this case, the power of reflected light from the surface R is dependent on the light arriving from each lighting source and is modelled by Lambert's cosine law:

$$R(\lambda) = a(\lambda) \sum_{i=1}^{N_i} (L_i(\lambda) \cos \theta_i) \quad (3)$$

where θ_i is the angle between the surface normal vector and the light arriving at the surface, L_i , for each light (see Figure 2 (c)) and $a(\lambda)$ is the albedo of the surface at wavelength λ .

3.2.3 Water-based Attenuation and Backscatter

As light travels through water it is attenuated according to the distance travelled. Equation 4 describes the relationship between light power at a source L_0 and destination L_d separated by distance d (Jaffe, 1990):

$$L_d(\lambda) = L_0(\lambda) e^{-b(\lambda)d} \quad (4)$$

where $b(\lambda)$ is the attenuation coefficient of the water at wavelength λ . In the underwater image formation model, light is attenuated both on its way from the light source to the imaged point and from the imaged point back up towards the camera. The light power arriving at the surface from each light source, L_i , is thus:

$$L_i = P_{\phi_i} e^{-b(\lambda)r_{li}} \quad (5)$$

where r_{li} is the range from light i to the imaged point. The light power arriving at the camera once reflected

from the surface is:

$$L_c = R(\lambda)e^{-b(\lambda)r_c} \quad (6)$$

where L_c is the light arriving at the camera corresponding to the surface point and r_c is the range from the camera to the point.

In addition to attenuation, the model also considers backscattered light. Figure 2 (d) illustrates the effect of backscattered light on images. Light coming from the two light sources is randomly scattered as it travels through the water medium in front of the camera. For all scattering events that occur with light passing through (but not in the same direction as) the exact line of sight between the surface point and the camera, some proportion of the scattered light will be reflected back towards camera. If we make the approximation that the proportion of scattered light ($\gamma(\lambda)$) has a uniform directional distribution, then the total quantity of backscattered light arriving at the camera ($B(\lambda)$) is given by Equation 7 (Jaffe, 1990):

$$B(\lambda) = \int_0^{r_c} \gamma(\lambda)L(r)e^{-b(\lambda)r} dr \quad (7)$$

where $L(r)$ is a function describing the total light arriving at each point of the camera line of sight as a function of the distance along the line of sight (r). In practise, we found that $L(r)$ could be approximated as being constant, and that the resulting magnitude of backscattered light could be computed by:

$$\begin{aligned} B(\lambda) &= \int_0^{r_c} \gamma(\lambda)L e^{-b(\lambda)r} dr_c \\ &= \frac{\gamma(\lambda)L}{b(\lambda)} \left[1 - e^{-b(\lambda)r_c} \right] \\ &= \frac{\beta(\lambda)}{b(\lambda)} \left[1 - e^{-b(\lambda)r_c} \right] \end{aligned} \quad (8)$$

where $\beta(\lambda) = \gamma(\lambda)L$ is a constant.

3.2.4 Camera Model

Light travelling through the lens of the camera undergoes a fade-out in intensity towards the corners of the image via the effect of vignetting (Kim and Pollefeys, 2008). Vignetting is caused primarily by the geometry of light passing through the lens and aperture of the camera; light passing in from greater angles to the principle axis of the camera is partially shaded by the aperture and sometimes by the lens housing. Vignetting can be summarised by:

$$E = C(\alpha)L \quad (9)$$

where L is the light arriving at the front of the lens, E is the light arriving at the camera image sensor (irradiance) and $C(\alpha)$ is a vignetting coefficient, modelled as a polynomial function of α , the angle between the light ray entering the camera and the principle axis of the camera frame (vector that crosses through the principle point in the image and the camera focal point):

$$C(\alpha) = 1 + C_{\alpha 2}\alpha^2 + C_{\alpha 4}\alpha^4 + C_{\alpha 6}\alpha^6 \quad (10)$$

where $C_{\alpha 2}$, $C_{\alpha 4}$ and $C_{\alpha 6}$ are polynomial coefficients. The last step in image formation occurs when light arriving at the image sensor of the camera is converted into an image intensity value I , via the sensor response function of the camera $f(\cdot)$:

$$I = f(kE) \quad (11)$$

where k is the exposure constant, typically proportionate to the shutter speed of the camera. The sensor response function $f(\cdot)$ can take a variety of forms, for example a gamma curve, and is typically controlled by camera firmware. A detailed discussion of sensor response functions can be found in (Grossberg and Nayar, 2004). In our model, we assume that the sensor response function is linear and normalised between zero and one, such that:

$$I = kE \quad (12)$$

For cameras with a non-linear response function, the method described in (Grossberg and Nayar, 2004) can be used apply an inverse of the response function $f(\cdot)$ that is normalised in this way.

3.2.5 Complete Image Formation Model

By combining Equations 1, 3, 5, 6, 7, 10 and 12 the complete image formation model is described by Equation 13:

$$I(\lambda) = k \left[C(\alpha) a(\lambda) \sum_{N_l}^{i=1} \left(P_{\phi_i} \cos \theta_i e^{-b(\lambda)(r_c+r_{li})} \right) + B(\lambda) \right] \quad (13)$$

The parameters r_c , r_{li} , ϕ_i , θ_i and α can all be measured from the physical geometry of the relative positions and pointing angles of the lights, camera and scene point which can easily be extracted from data provided by the underwater structure-from-motion pipeline described in Section 3.1. We assume the exposure constant k is known for each image from the recorded shutter speed of the camera. The parameters C_{α_2} , C_{α_4} and C_{α_6} are unknown coefficients of the camera used and the parameters $b(\lambda)$ and $\beta(\lambda)$ are unknown coefficients of the water column.

Given a means to estimating the unknown parameters described above, an inverse version of Equation 13 can be used to compute the albedo of a scene point from the measured intensity of the object within an image:

$$a(\lambda) = \left[\frac{I(\lambda)}{k} - B(\lambda) \right] \frac{1}{K} \quad (14)$$

$$K = C(\alpha) \sum_{N_l}^{i=1} \left(P_{\phi_i} \cos \theta_i e^{-b(\lambda)(r_c+r_{li})} \right)$$

By setting the light source power $P_{0,i} = 1$ (a unit value) for each of the light sources, the calculated albedo is relative to the power of the light source at the specified wavelength of the channel and the corrected image intensities correspond to the intensity of the object imaged under the spectral colour of the light source. A further processing step is required for calculating the albedo relative to a white light source (using knowledge of the light spectral power) and is discussed below in Section 3.4.

3.3 Estimation of Image Formation Parameters

In this section we develop an approach to estimating the unknown underwater image formation model parameters discussed above. Consider Equation 13; at first it seems necessary to provide a collection of measured image point intensities and known corresponding object albedos in order to compute the unknown coefficients of the water column or camera response/vignetting. In this section we develop a method for estimating these parameters using only the measured image point intensities of a collection of surface points. Our method simultaneously estimates the albedos of the scene points along with the camera/water parameters by using multiple co-registered observations of each point from different image perspectives. Our approach computes a maximum-likelihood estimate, using non-linear least squares and Levenberg-Marquardt optimisation of the unknown parameters using a small subset of the total number of scene points available on the 3D terrain surface provided by structure-from-motion. Once the camera and water column parameters have been computed, the albedo of any point on the terrain surface can be computed using Equation 14. We estimate different sets of parameters for each of the red, green and blue channels collected by the camera.

3.3.1 Estimated Parameter and Observation Vectors

The parameter estimation procedure begins by selecting a set of small triangular surface elements lying on the terrain that have each been observed multiple times and co-registered across the corresponding series of images (see Figure 3). A subset of N triangles across the whole terrain surface is selected randomly, rather than considering every single triangular face. For each image channel, the estimated parameter vector \mathbf{x} is composed of the albedo of each of these triangles along with the attenuation and backscatter coefficients and camera vignetting parameters, all at the corresponding wavelength of the channel:

$$\mathbf{x} = [b, \beta, C_{\alpha 2}, C_{\alpha 4}, C_{\alpha 6}, a_1, a_2, \dots, a_N]^T \quad (15)$$

For each triangle, the positions of the 3D vertex points are projected into the camera reference frame for each of the captured images and used to select the set of images from which the triangle was seen. For each corresponding image, the midpoint of the triangle is used to compute the range from the triangle to

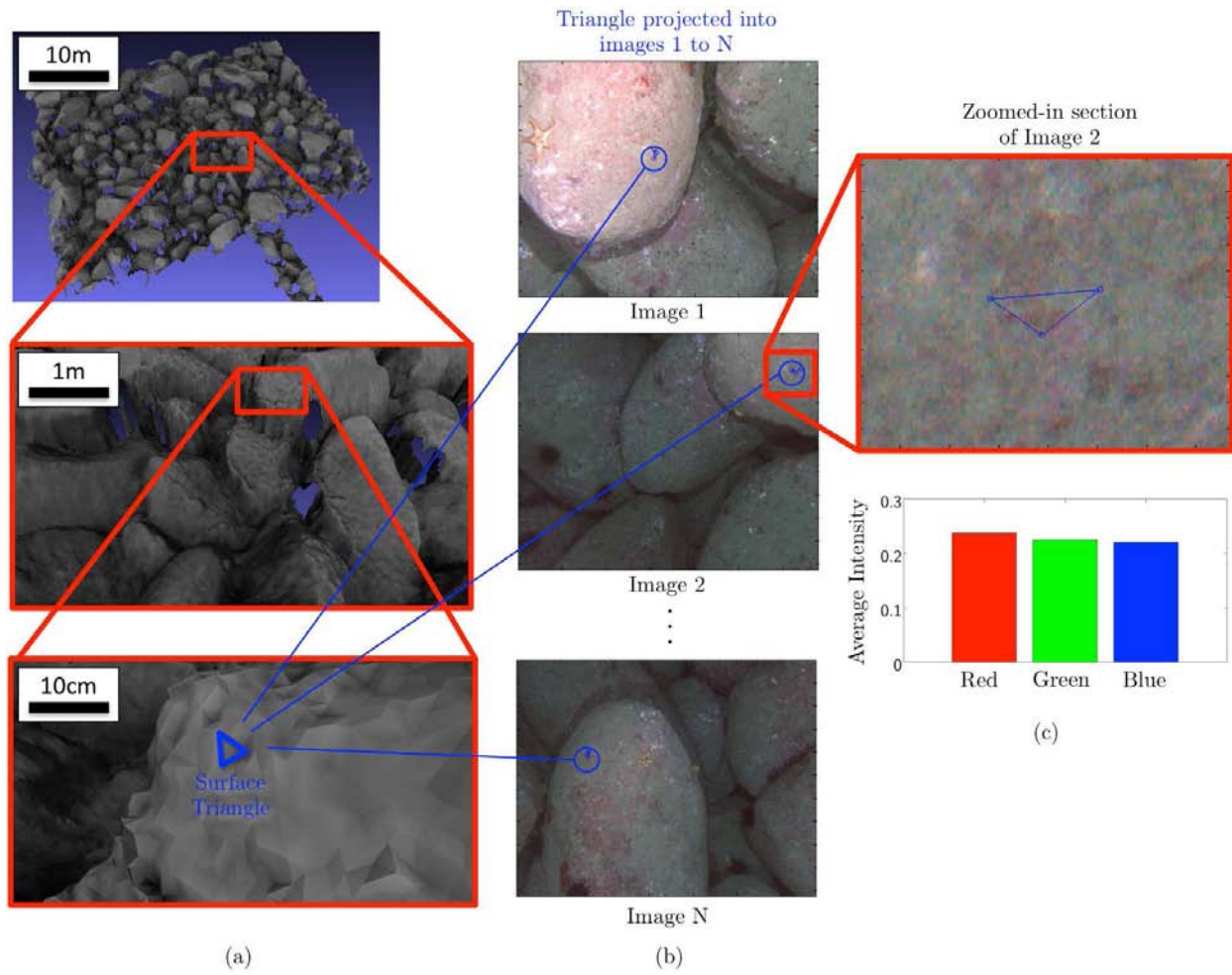


Figure 3: Extracting image intensity measurements from structure-from-motion data for image formation model parameter estimation: (a) Recursively zoomed-in views of 3D terrain surface model with selected scene triangle. (b) Example images that captured the scene triangle with triangle reprojected into each image using structure-from-motion computed poses. (c) highlighted section of image 2 showing reprojected triangle and bar graph of average red, green and blue intensities corresponding to the pixels inside the triangle.

the camera and to each of the lights. Using the vector from each light to the triangle midpoint and the triangle normal vector (both referenced in the camera frame), the angles ϕ and θ are computed for each light source. The intensity of the pixels lying within the reprojected triangles are averaged for each of the red, green and blue channels (see Figure 3). Since values are averaged over each triangle, our method assumes the case that the triangular surface regions are small with respect to the distance from the camera (in our case approximately 3cm across the longest axis whereas images are captured at a range of approximately 1.5 to 5m from the surface).

For a given channel, the observation vector \mathbf{z} is composed of the average measured image intensities (I) for the reprojected triangle in each image of each scene triangle:

$$\mathbf{z} = [I_1^1, I_1^2, \dots, I_1^{M_1}, I_2^1, I_2^2, \dots, I_2^{M_2}, \dots, I_N^1, I_N^2, \dots, I_N^{M_N}]^T \quad (16)$$

where for each I the subscript represents the scene triangle number (1 to N) and the superscript represents the image number for a given scene triangle (1 to M_i), where M_i is the total number of images of the i th triangle on the 3D terrain surface. The relationship between an observation I_i^j (i.e. the j th observation of the i th scene triangle) and the estimated parameters is described by the function $h_{i,j}(\cdot)$:

$$I_i^j = h_{i,j}(b, \beta, C_{\alpha 2}, C_{\alpha 4}, C_{\alpha 6}, a_i) \quad (17)$$

which is identical in form to Equation 13 (note that the terms r_c , r_{l1} , r_{l2} , ϕ , θ , α and k are all assumed to be known and thus inherent to $h_{i,j}(\cdot)$ itself). Therefore the total observation function $\mathbf{h}(\cdot)$ is defined by the concatenation of observation functions for all observations of all scene points and is described by the relationship $\mathbf{z} = \mathbf{h}(\mathbf{x})$.

3.3.2 Non-linear Least Squares and Levenberg-Marquardt

The aim of the parameter estimation procedure is to estimate a parameter vector $\hat{\mathbf{x}}$ which minimises the weighted non-linear least square cost function:

$$\hat{\mathbf{x}} = \operatorname{argmin} \frac{1}{2} (\mathbf{z} - \mathbf{h}(\hat{\mathbf{x}}))^T \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \mathbf{h}(\hat{\mathbf{x}})) \quad (18)$$

where $\boldsymbol{\Sigma}$ is a diagonal covariance matrix, representing the modelled noise in the measured image intensities and $\mathbf{h}(\hat{\mathbf{x}})$ is the predicted value of the observation vector given a parameter estimate $\hat{\mathbf{x}}$.

Our solution procedure uses a Levenberg-Marquardt optimisation (Marquardt, 1963) to iteratively converge on the final estimate using an initial parameter guess $\hat{\mathbf{x}}_0$. At each step of the iteration, an updated parameter estimate $\hat{\mathbf{x}}_{\text{new}} = \hat{\mathbf{x}}_{\text{prev}} + \delta\mathbf{x}$ is computed from the previous parameter estimate $\hat{\mathbf{x}}_{\text{prev}}$ and update vector $\delta\mathbf{x}$, computed by solving the linear system in Equation 19:

$$[\mathbf{Y} + \omega \mathbf{diag}(\mathbf{Y})] \delta\mathbf{x} = \mathbf{y} \quad (19)$$

where $\mathbf{diag}(\mathbf{Y})$ is the diagonal of the matrix \mathbf{Y} , ω is the Levenberg-Marquardt adaptive damping factor and:

$$\mathbf{Y} = \nabla \mathbf{H}^T \boldsymbol{\Sigma}^{-1} \nabla \mathbf{H} \quad (20)$$

$$\mathbf{y} = \nabla \mathbf{H}^T \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \mathbf{h}(\mathbf{x})) \quad (21)$$

where $\nabla \mathbf{H}$ is the Jacobian of $\mathbf{h}(\mathbf{x})$ computed at the current estimate of $\hat{\mathbf{x}}$, and \mathbf{Y} , \mathbf{y} are commonly referred to as the information matrix and vector respectively. Sparse matrix methods are used to store and perform operations on \mathbf{Y} using a Cholesky factorisation in column-compressed form (Davis, 2006).

3.3.3 Removing Outliers

Due to a variety of factors, not all of the measured image intensities in the observation vector \mathbf{z} fit well to the image formation model described in Section 3.2. Factors that cause a large discrepancy between the mea-

sured image intensity I_i^j and predicted image intensity $h_{i,j}(\hat{\mathbf{x}})$ include image noise, errors in the reconstructed terrain model (incorrectly recovered surface angle), camera pose errors (surface triangle misregistration) and object surfaces with largely specular reflection. In order to remove these effects, an automatic outlier detection scheme is built into the parameter estimation procedure that rejects observations and scene triangles from \mathbf{z} and \mathbf{x} that violate the model assumptions, detected through the associated residual vector $\rho = \mathbf{z} - \mathbf{h}(\hat{\mathbf{x}})$ during the estimation cycle.

During each step of the estimation cycle, the average residual is computed from the residual vector ρ . Observations corresponding to a residual with an absolute value greater than three times the average residual are removed from the observation vector, and when two or more outliers are discovered from a given scene triangle, the triangle itself is removed from the parameter vector.

3.3.4 Non-linear Least Squares Procedure

The overall parameter estimation procedure is shown in Algorithm 1.

Algorithm 1: Non-linear least squares procedure based on Levenberg-Marquardt optimisation for estimating underwater image formation parameters and scene triangle albedos using the measured intensities in co-registered image points.

Input: $\mathbf{z}, \hat{\mathbf{x}}_0$

Output: $\hat{\mathbf{x}}$

Set initial guess $\hat{\mathbf{x}} = \hat{\mathbf{x}}_0$; $tol = 1 \times 10^{-9}$; $norm(\delta\mathbf{x}) = 1$;

while $norm(\delta\mathbf{x}) > tol$ **do**

 Compute residual vector $\rho = \mathbf{z} - \mathbf{h}(\mathbf{x})$;

for $i = 1$ to m **do**

 | if $abs(\rho_i)$ is greater than three times average residual, remove observation;

end

 Compute $\nabla\mathbf{H}, \mathbf{Y}$ and \mathbf{y} ;

 Solve Equation 19 for $\delta\mathbf{x}$ using Cholesky factorisation and triangular system solving;

 Update parameter estimate $\hat{\mathbf{x}} = \hat{\mathbf{x}} + \delta\mathbf{x}$;

 Adjust ω based on convergence criteria (see (Marquardt, 1963));

end

In practice, the estimation procedure exhibited good convergence properties for a range of different initial parameter guesses. The initial parameter guess $\hat{\mathbf{x}}_0$ was set with the parameters $b, \beta, C_{\alpha 2}, C_{\alpha 4}, C_{\alpha 6} = 0$ and the scene triangles albedos a_1, a_2, \dots, a_N as the average of the measured intensities across the corresponding images relating to that triangle (i.e. $a_N = \frac{1}{M_N} \sum_{j=1}^{M_N} I_N^j$). We found, for the observations made in our

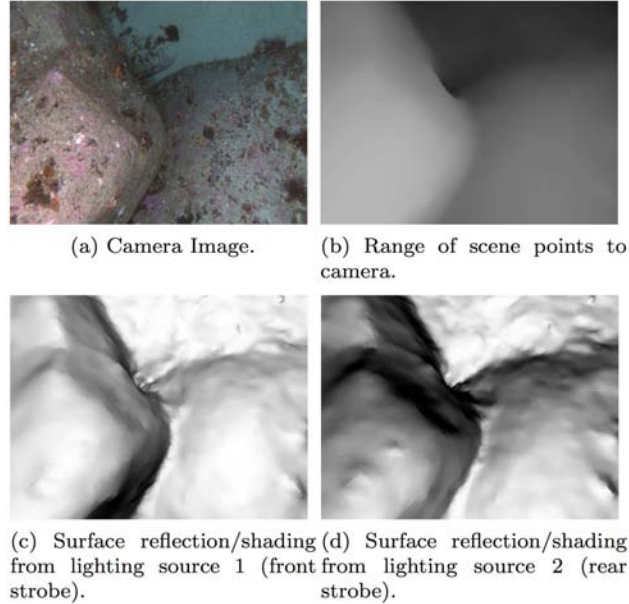


Figure 4: Rendering of image formation model parameters in the image correction process: (a) A colour image captured by the AUV. (b) Range-image of the scene where the intensity represents the range from the camera to the scene surface at each pixel. (c, d) Surface reflection/shading images where the intensity represents the $\cos\theta$ value (cosine of angle between the light pointing vector and surface normal vector) for each pixel, where (c) is for reflection with respect to a front-mounted strobe and (d) is for a rear-mounted strobe.

experimental datasets, that we could set the initial guess of the attenuation and backscatter coefficients at any value between $b = 0$, $\beta = 0$ and a reasonable guess for standard seawater (i.e. $b = 0.2 - 0.5$, $\beta = 0.02 - 0.06$ (Mobley, 1994)) and still achieve the same stable convergence in the estimation of these parameters using the Levenberg-Marquardt optimisation.

3.4 Re-rendering True Colour Images

3.4.1 Re-rendering Images

Once the image formation parameters have been estimated, Equation 14 is used to recover the albedo of each pixel in each image. Values for the camera-to-surface and light-to-surface distances and angles (i.e. r_c , r_{l1} , r_{l2} , ϕ , θ) are computed at each pixel by linearly interpolating between the values at the nodes of the triangles in the terrain surface model, projected into the camera's frame of reference. Figure 4 illustrates the parameter rendering process for some of the model parameters. Once the albedo at each pixel has been recovered, a set

of ‘true-colour’ images is reconstructed. Images are applied through an air-based and vignetting-free image formation model which assumes the scene is illuminated from directly above and captured with a constant exposure time. For each pixel, the air-based image intensity $I_{air}(\lambda)$ is computed using Equation 22:

$$I_{air}(\lambda) = a(\lambda) \cos \theta_z \quad (22)$$

where θ_z is the angle between the surface normal vector and the vertical vector $[0, 0, 1]^T$, representing sunlight from directly overhead.

3.4.2 Accounting for Camera Spectral Response and Artificial Lighting Spectrum

When applied over the red, green and blue colour channels, Equation 22 results in an image where the intensity of colours represents that seen by the camera in air and illuminated from above, but with an illumination colour based on that provided by the artificial lighting carried by the AUV. In the case where knowledge of the camera spectral response function and illumination spectrum are available, these functions can be used to produce images as seen by a white light source. This processing step is useful for example in situations in which images from different dives of a single area, captured using differently coloured light sources are to be compared.

The camera spectral response function $g(\lambda)$ represents, for a given colour channel, the sensitivity of the camera’s sensor to light at a particular wavelength λ . The illumination source spectrum function $j(\lambda)$ represents the power of the light output over different wavelengths. When these functions are known (typically provided by camera/light manufacturers or can be determined using simple laboratory procedures (Finlayson et al., 1998)), the estimated response of the camera to a white target, under the lights, in each colour channel can be estimated. For the red, green and blue colour channels, the expected white object responses are:

$$T(red) = \int g_{red}(\lambda)j(\lambda)d\lambda \quad (23)$$

$$T(green) = \int g_{green}(\lambda)j(\lambda)d\lambda \quad (24)$$

$$T(blue) = \int g_{blue}(\lambda)j(\lambda)d\lambda \quad (25)$$

where T_{red} , T_{green} and T_{blue} are the white target expected image intensities and $g_{red}(\lambda)$, $g_{green}(\lambda)$ and $g_{blue}(\lambda)$ are the camera response functions for the red, green and blue channels respectively. Alternatively, T_{red} , T_{green} and T_{blue} can be determined by imaging a spectrally-white reflection panel using the camera and light source. The corrected image intensities in Equation 22 can then be normalised to a white light source/balanced camera via the equations:

$$I_{air,white}(red) = \frac{\frac{1}{T(red)}}{\max(\frac{1}{T(red)}, \frac{1}{T(green)}, \frac{1}{T(blue)})} I_{air}(red) \quad (26)$$

$$I_{air,white}(green) = \frac{\frac{1}{T(green)}}{\max(\frac{1}{T(red)}, \frac{1}{T(green)}, \frac{1}{T(blue)})} I_{air}(green) \quad (27)$$

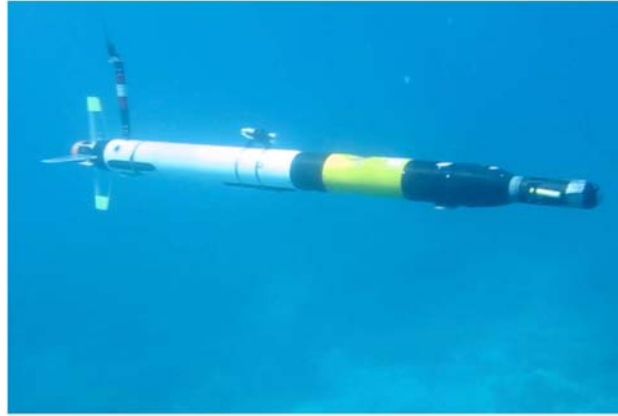
$$I_{air,white}(blue) = \frac{\frac{1}{T(blue)}}{\max(\frac{1}{T(red)}, \frac{1}{T(green)}, \frac{1}{T(blue)})} I_{air}(blue) \quad (28)$$

to produce an image that would be obtained for a white-coloured illumination and a camera with equal sensitivities in the red, green and blue channels respectively.

4 Experimental Setup

In order to evaluate the proposed image correction technique, we applied our algorithm to imagery collected during two different AUV dives using two different AUV platforms (Iver and Sirius, see Figure 5). Both platforms carried a sensor suite consisting of a stereo-camera system, depth sensor, attitude heading and reference system and Global Positioning Systems (for surface operation) which were logged and post-processed to produce 3D, geo-referenced structure-from-motion models of the underwater terrain over which they operated. The specifications of the sensors carried in both experiments are shown in Table 1.

Both AUV systems fielded a stereo camera rig that was comprised of a colour Bayer-sensor camera and a monochrome camera (all colour processing was applied to the colour images only). Images were captured in a raw format, which was later de-mosaicked to three colour channels at a resolution of 1360-by-1024 pixels



(a) Iver AUV



(b) Sirius AUV

Figure 5: Experimental setup: AUV Iver and AUV Sirius.

with a depth of 12-bits and with constant exposure parameters. Artificial light was provided by two strobes on each platform that were placed at the front and rear of the AUVs. Images were captured in a downwards orientation while each AUV moved back-and-forth across the terrain, producing many overlapping images.

For each dive, we compared images produced by our correction scheme to the original, unprocessed images and to images processed using a grayworld algorithm (see (Bryson et al., 2012; Buchsbaum, 1980)). The grayworld algorithm computes a scale and offset for each of the red, green and blue colour channels that is applied across all images such that the resulting distribution of intensities in each channel matches a specified mean and variance. When the desired mean and variance is set to be equal in each channel, the colour channels are well matched, under the assumption that, on average, the colour of objects in all of the images is gray. The grayworld correction is frequently applied in underwater images to balance the blue hue

shift which is often observed due to attenuation. The specified mean and variance of the image signals used in the grayworld algorithm were 0.5 and 0.16^2 respectively.

4.1 Clovelly Validation Dataset (Iver)

Imagery was collected using the Iver AUV at Clovelly beach, Sydney, Australia in a validation dataset to quantitatively assess the accuracy of the colour correction technique. The dive site was close to the shore and within a relatively sheltered area, which allowed for a diver to place colour reference targets in the mission area. Imagery was collected over an area of approximately 10-by-10m in a rubble dominated bottom type in about 3-4m of water depth and at a range of platform altitudes from approximately 1 to 4m. A Macbeth colour reference chart was placed in the mission area, which provided 24 different colour patches, with known spectral reflectance data. The dive was performed at night such that the ambient light had a negligible effect on the seafloor. In addition to the underwater images, images were taken of the Macbeth colour chart in air by holding the platform out of the water using the same camera and lights, to provide air reference images against which the corrected underwater images could be assessed.

After the data was collected, the SLAM/meshing software pipeline described in (Johnson-Roberson et al., 2013; Johnson-Roberson et al., 2010; Mahon et al., 2008) was used to build 3D photo-textured models of the seafloor area surrounding the dive site, including the colour reference targets, using approximately 700 stereo image pairs. A random subset of 1000 faces of the resulting 3D terrain surface model were used following the method outlined in Section 3.3 to estimate the image formation model parameters. The selected faces were examined to ensure that none corresponded to the Macbeth colour chart itself to ensure that the algorithm would only be provided with naturally-occurring surfaces. The estimated parameters were used in Equations 14 and 22 with the original unprocessed underwater images to produce corrected images. Grayworld-corrected images were produced using the statistics across all 700 colour images.

For each set of images (uncorrected, grayworld-corrected and image model-corrected), 24 patches from each of 97 of the images (those in which the Macbeth colourboard was visible) were extracted corresponding to each of the 24 Macbeth colour panels. In order to compare the colour intensities of the extracted panels in

each set of images to those in the in-air reference image, the image intensities were normalised by using a linear transformation based on the average intensities of the brightest panel (Macbeth white panel, ID 18) and the darkest panel (Macbeth black panel, ID 23):

$$I_{norm,j}(\lambda) = mI_j(\lambda) + o \quad (29)$$

where

$$m = \frac{1}{\mu_{white} - \mu_{black}} \quad (30)$$

$$o = \frac{\mu_{black}}{\mu_{white} - \mu_{black}} \quad (31)$$

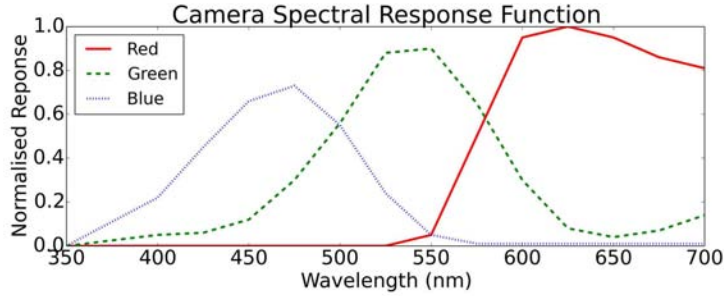
$$\mu_{white} = \frac{1}{3} \sum_{\lambda=(R,G,B)} \frac{1}{N_{image}} \sum_{j=1}^{j=1} I_j^{white}(\lambda) \quad (32)$$

$$\mu_{black} = \frac{1}{3} \sum_{\lambda=(R,G,B)} \frac{1}{N_{image}} \sum_{j=1}^{j=1} I_j^{black}(\lambda) \quad (33)$$

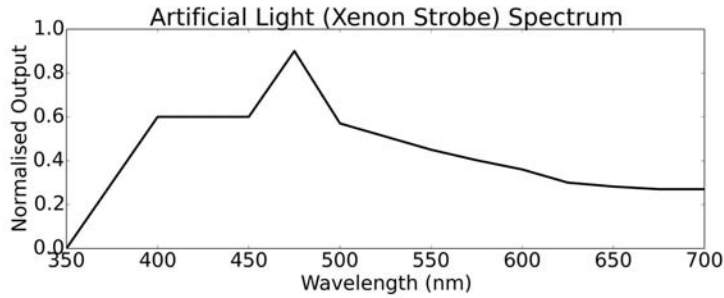
where, for a given set of images, $I_{norm,j}$ is the normalised image intensity of the j^{th} image, μ_{white} is the average image intensity of the white panel and μ_{black} is the average image intensity of the black panel across all N_{image} images and three colour channels. The normalised intensities therefore maintained the colour balance present in each set of images, but scaled the magnitudes of absolute intensity to be equal, such that comparison was possible.

4.2 Tasmania Urchin Barrens Dataset (Sirius)

Imagery collected by the Sirius AUV over an urchin barren off the coast of St Helens, Tasmania, Australia was used to assess the performance of the colour correction algorithms both quantitatively and qualitatively. Approximately 6300 images were collected over a 25-by-15m area that was made up predominantly of large boulders, in water depths of approximately 30m. Owing to the rugous nature of the terrain, the AUV's altitude varied from about 1 to 3m during the dive. The image formation model parameters in Equation 15 were estimated using our technique. Extracted raw image data was used from 1000 randomly selected



(a) Camera Spectral Response Functions



(b) Artificial Strobe Lighting Spectrum (Xenon strobe, AUV Sirius)

Figure 6: Camera spectral response functions (for each colour channel) and artificial strobe lighting spectrum (taken from manufacturer’s data).

faces for the parameter estimation phase of the algorithm and images from the entire dive were corrected using the image model-correction scheme and grayworld correction algorithm. The resulting images were used to photo-texture a 3D model of the terrain (generated through the SLAM/meshing process described in (Johnson-Roberson et al., 2013; Johnson-Roberson et al., 2010; Mahon et al., 2008)) and orthographic projections of this model were used to render overhead imagery mosaics.

No colour reference data was available for this dive. Therefore, to assess the performance of the colour correction scheme we computed the variations in colours in each terrain model face based on all of the images that observed a given face. The image sets that were produced by the different methods (uncorrected, grayworld-corrected and model-corrected) were gain-normalised, such that variance measures were comparable. For each set of images, a gain was applied to the intensities such that the standard deviation of all intensities in all images was equal to one. As in Section 4.1, the same gain value was applied to all of the red, green and blue image channels so as to preserve the colour balance in each method. Once the images were normalised,

for each terrain face, we computed all of the images that saw that face (using the structure-from-motion relationships, typically 10-100 images) and reprojected the face coordinates into each image. The red, green and blue intensities of the reprojected patches were extracted. The standard deviation of extracted colours was computed, and these standard deviation values averaged over all faces in the terrain model.

The spectral response functions of the camera and xenon strobe light sources were ascertained from manufacturer's data (see Figure 6) and used to compute white-balanced scaling parameters using Equations 26 to 28. The image model corrected images were then scaled using these equations and the resulting images used to create an imagery mosaic for which the colours represented an approximation of the scene in air and illuminated by white lights from directly overhead.

5 Results

5.1 Clovelly Dataset

Figure 7 shows example images from various perspectives from the Clovelly dataset of the unprocessed underwater images, grayworld-corrected images and the image formation model-corrected images. The uncorrected images (Figure 7 (a)) exhibit a strong blue hue, owing to the colour attenuation of the water. The grayworld images (Figure 7 (b)) exhibit a colour balance that is much closer to what would be observed in air, but still with a slight blue hue and, like the uncorrected images, exhibit large variations in intensity and colour shifts depending on the range to the camera and position with respect to the lights. The image model-corrected images (Figure 7 (c)) exhibit colours which represent the appearance of the scene in air, under that illumination of the light source on the AUV (a slight green hue). The variation in colour and intensity in the images with changes in perspective are minimal, owing to the model correction, and as would be expected if the scene were to be imaged in air and accounting for the camera vignetting and light spatial distribution.

Figure 8 illustrates rendered Macbeth colour charts using the average colours present across the 97 images

in which the Macbeth colour chart was in-frame. Figure 8 (a) shows the colours present in the uncorrected images, which have a strong blue-green colour cast. Figure 8 (b), shows the colours for the grayworld-corrected images, Figure 8 (c) shows the colours for the image model-corrected images and Figure 8 (d) shows the colours from an image captured in air. The grayworld image colours are more closely matched to the in-air colours than those of the unprocessed images, however the image model-corrected colours most closely approximate the in-air images.

Figure 9 illustrates the range of colours present in each image set for one of the Macbeth colour panels (Neutral 8, corresponding to a bright, gray colour) plotted as a function of the range of the colour chart to the camera in each image. The uncorrected and grayworld corrected image colours vary significantly and are strongly correlated to range, whereas the image-model corrected image colours exhibit almost no correlation to range and are closer to the actual gray of the panel. The panel colours do display small shifts away from gray at the very closest and furthest ranges; this is indicative that the parameters estimated for the image formation model probably do not provide a good estimate of image intensities beyond the ranges of which images were actually observed during the fitting process (i.e. the model may have difficulty extrapolating beyond range values observed in our dataset).

Figure 10 shows the means and standard deviations of the normalised colour errors for each of the uncorrected, grayworld-corrected and image model-corrected image sets when compared to the colours present in the in-air images. The uncorrected images display both a large mean offset and standard deviation of colour intensities. The grayworld-corrected images display a reduced mean colour error, but an equivalent level of error in the standard deviation of colours from the in-air images. The image model-corrected images display both a reduced mean error and reduced error standard deviation when compared to the other approaches, demonstrating the ability of the method to reproduce both accurate and consistent colour.

5.2 Tasmania Urchin Barrens Dataset

Figure 11 shows example images from various perspectives from the Tasmania urchin barrens dataset of the unprocessed underwater images, grayworld-corrected images and the image formation model-corrected

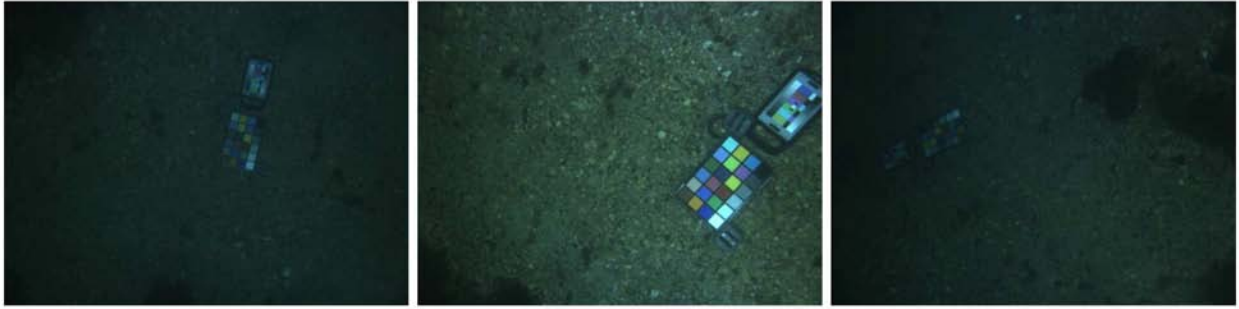
images. The uncorrected and grayworld-corrected images exhibit changes in colour and intensity across the regions of each image, owing to the range of object depths visible within a single scene, whereas in the model-corrected images these variations are greatly mitigated.

Figure 12 shows imagery mosaics of the study site constructed using the images from each of the processing schemes. For both the uncorrected images and grayworld-corrected images, the mosaics exhibit sharp inconsistencies in colour and intensity across co-located faces owing to the different images used to texture each face, which are captured from varying perspectives and ranges. Most prominent are the horizontal strips of light and dark that correspond to the tracklines of the AUV as it passes over the terrain. The terrain at this site was sloped along the direction of the tracklines of the AUV; the use of a forward looking altitude-hold system on the AUV meant that the vehicle did not follow a consistent altitude when heading in different directions during multiple, overlapping passes (i.e. heading up the slope versus heading down the slope). The resulting trajectories meant that images were captured across different ranges to the surface. In contrast to Figures 12 (a) and (b), the mosaic produced using the model corrected images in Figure 12 (c) does not contain this artifact; the image colours and intensities do not vary with imaging perspective and are consistent across the terrain.

Figure 13 shows a comparison of the average within-face image colour standard deviations for each of the processing schemes. As in other results, the image model-corrected face textures display a reduced variation in colour intensity with perspective changes when compared to the uncorrected and grayworld corrected sets. Since the grayworld correction scheme applies only a single gain and offset value to an image set, the average face variation is equivalent to that of the uncorrected images when normalised to the same output distribution, as is shown in this figure. The variation in face colours for the image model correction scheme is lowest owing to the higher consistency in multiple images of the same surface. This performance metric essentially represents the average ratio of image intensity variations within a single terrain model face to the variation of all colours in all faces. Since the total variation in intensities in the uncorrected and grayworld-corrected sets is already high, owing to the inconsistencies in changes in depth, this performance measure is in fact unfairly biased against the image model-correction scheme, which exhibits lower overall variations.

Despite this, the correction scheme still outperforms the other methods.

Although the images and mosaic textures present in the model-corrected images are highly consistent, the images still display a blue-green colour cast corresponding to the illumination colour of the artificial strobe lights carried on the AUV. Figures 14 and 15 compare example images and imagery mosaics before and after the images are adjusted to a white light source/balanced camera by applying Equations 26 to 28, which use the information about the camera colour sensitivity and light spectrum in Figure 6. The resulting images display colours that closely approximate the appearance of the scene as if imaged in-air and with white lights, illuminating the scene from directly above.



(a) Uncorrected underwater images.



(b) Grayworld correction scheme



(c) Underwater image formation model correction scheme

Figure 7: Example images at varying depths of the uncorrected, grayworld-corrected and image model-corrected images (Clovelly). The uncorrected and grayworld-corrected images exhibit intensity and colour shifts with varying imaging perspective, whereas the image-model corrected images exhibit a consistent colour cast, representing the appearance of the scene in-air and under the illumination colour of the lights carried by the AUV.



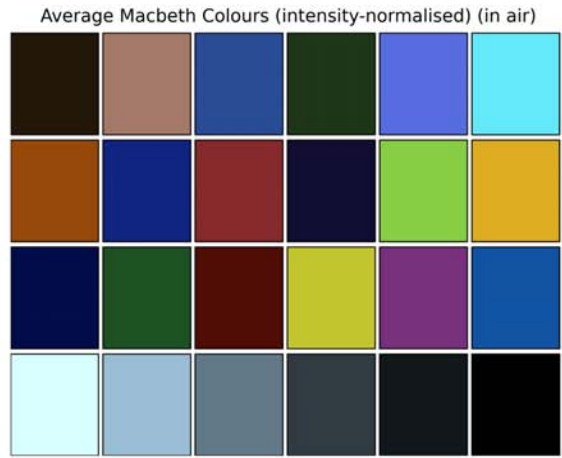
(a) Uncorrected underwater images.



(b) Grayworld correction scheme



(c) Underwater image formation model correction scheme



(d) Images in air

Figure 8: Rendered Macbeth colour panels using image data taken from raw underwater images, underwater images corrected using the grayworld and image model correction schemes, and images in air.

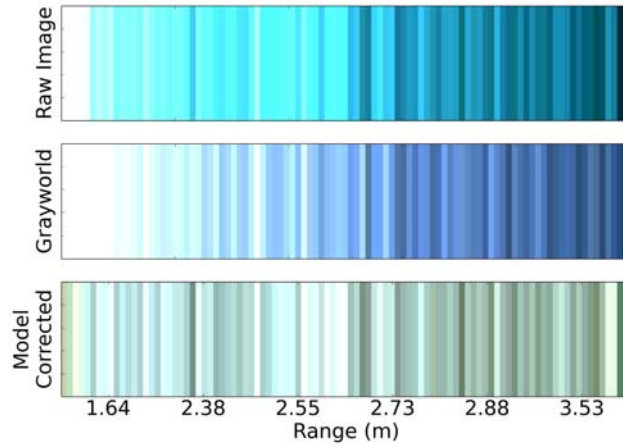


Figure 9: Variations in the appearance of grayscale Macbeth panel 19 (Neutral 8) vs. range of the panel to the camera for three image types: (i) raw, uncorrected images, (ii) images corrected using a grayworld correction scheme, and (iii) images corrected using our underwater image formation model correction scheme.

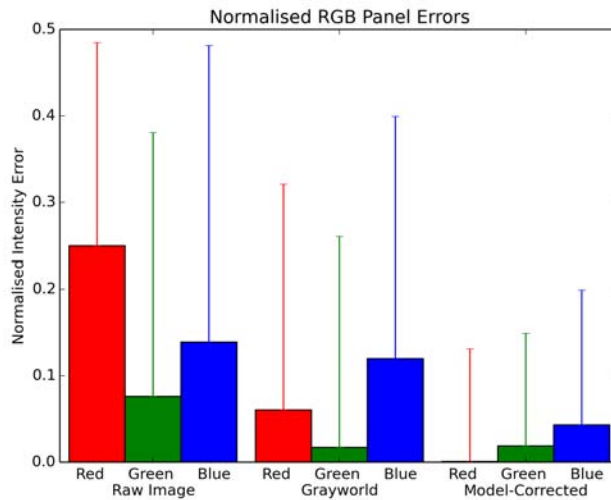


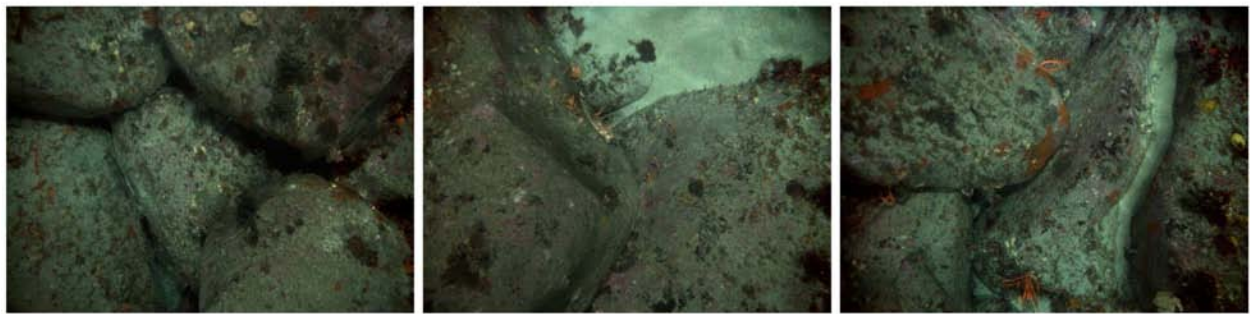
Figure 10: Comparison of the red, green and blue channel normalised intensity errors for the uncorrected images, grayworld-corrected images and image formation model-corrected images (errors measured from in-air reference image). The solid bars represent the absolute value of the mean error whereas the error bars represent the standard deviation of intensity values (from all images in the set) around this mean error.



(a) Uncorrected underwater images.

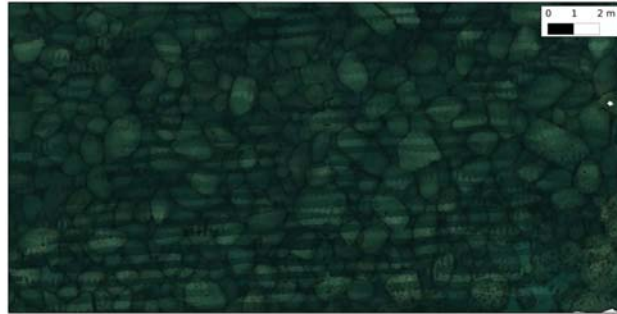


(b) Grayworld correction scheme

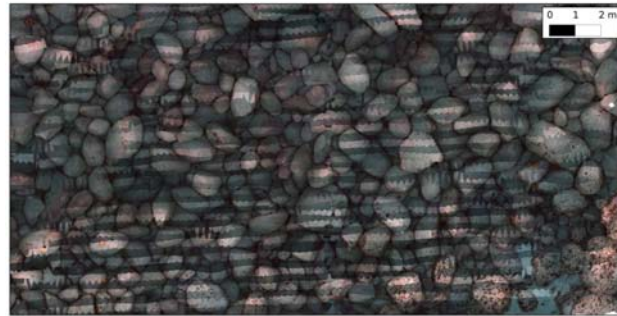


(c) Underwater image formation model correction scheme

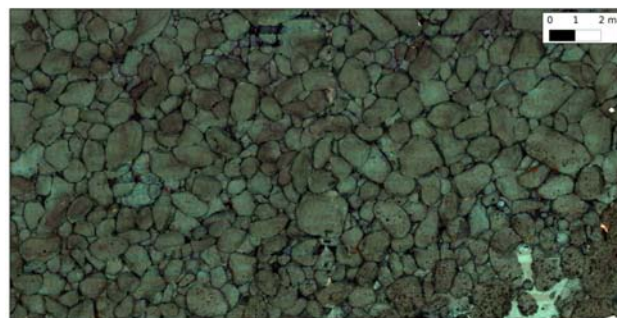
Figure 11: Example images at varying depths of the uncorrected, grayworld-corrected and image model-corrected images (Tasmania).



(a) Uncorrected underwater images.



(b) Grayworld correction scheme



(c) Underwater image formation model correction scheme

Figure 12: Imagery mosaic of Urchin barrens using image data taken from raw underwater images, underwater images corrected using the grayworld and image model correction schemes.

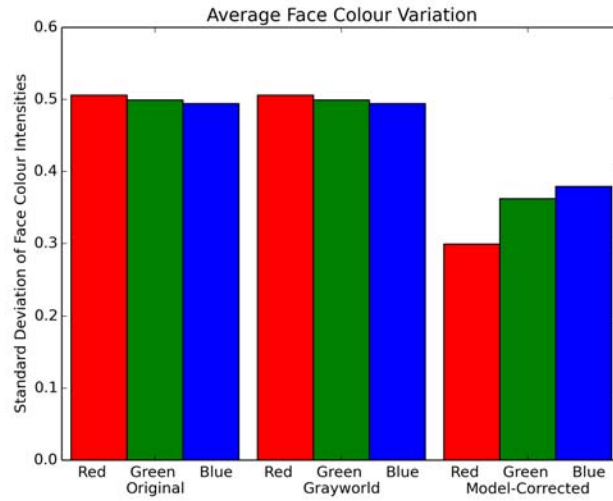
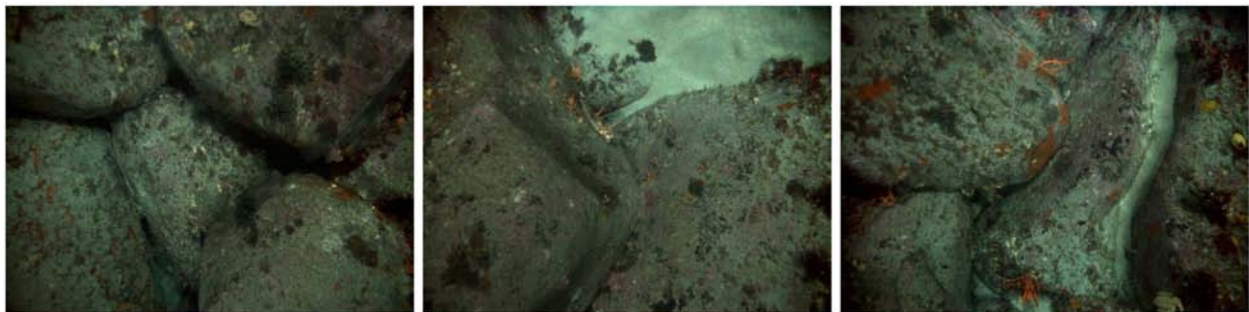


Figure 13: Average variation of colours in each face based on image projections for the uncorrected, grayworld corrected and image model corrected images. The intensity of all image sets were normalised using a gain parameter that resulted in the range of all colours in a given image set to have a unit standard deviation for each channel. The variation in face colours for the image model correction scheme is lowest owing to the higher consistency in multiple images of the same surface.

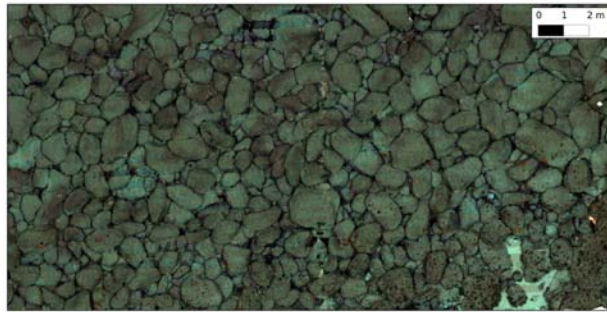


(a) Underwater image formation model correction scheme

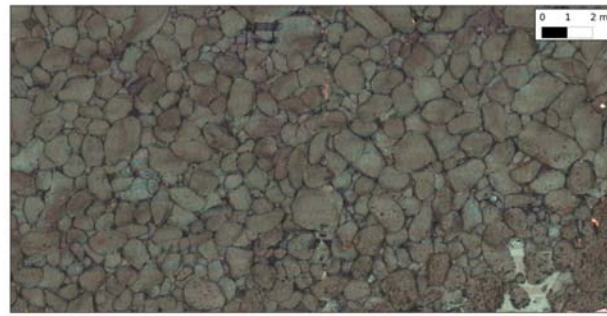


(b) Underwater image formation model correction scheme with camera/strobe spectral processing.

Figure 14: Example image model-corrected images before and after accounting for camera/lights spectral response function data to estimate the ‘normal’ colour of images (appearance under white lights with colour balanced camera).



(a) Underwater image formation model correction scheme



(b) Underwater image formation model correction scheme with camera/strobe spectral processing.

Figure 15: Imagery mosaic of Urchin barrens by adjusting colour gains of image model corrected imagery according to the known camera response and strobe lighting spectra.

6 Discussion

6.1 Advantages and Potential Uses of Our Method

The colour corrected underwater images produced by our method allow for colour to be effectively used in a quantitative fashion in which consistent colour measurements are required. This has potential implication for applications such as automatic image-based classification of benthic habitats and organisms by providing consistent colours as features for automatic classification based on supervised machine learning. Most current approaches to underwater image classification either avoid the use of colour as an explicit feature or utilise texture and other features (Beijbom et al., 2012; Soriano et al., 2001) due to the inconsistencies present when imaging from varying perspectives. The ability to generate consistent colours that are independent of the artificial lighting source provides the ability to quantitatively compare imagery collected over multiple repeated dives using different AUV platforms, which is an important part of long term marine habitat monitoring.

Although existing approaches using in-situ calibration setups (such as colourboards) also allow for this normalisation, our approach has the benefit that this infrastructure is not required. This significantly reduces the complexity of gathering data in sensitive or deep habitats in which manual placement of colourboards is difficult. The method also allows us to correctly process existing historical imagery datasets in which no colour calibration infrastructure was used during data collection.

6.2 Limitations

Our processing technique requires 3D structural information of the imaged terrain, referenced within the image data, to provide the necessary inputs to the image formation model. In cases in which this data is not available or cannot be ascertained from images (i.e. use of a monocular camera where there is no overlap between subsequent images), our technique cannot be used. Additionally, our technique is potentially susceptible to errors in the structure-from-motion processing. In practise, we found that the outlier rejection technique described in Section 3.3.3 was effective at removing these artifacts during the fitting process,

however these errors were still occasionally present in the corrected images themselves. The main source of error was due to incorrect recovery of terrain surface for complex structures at fine-scales that contained high-frequency variations in shading. Improved methods for recovering fine-scale terrain features (such as dense stereo processing) are likely to improve on this effect, however further investigation is outside of the scope of this paper.

Underwater images have inherently low signal-to-noise ratio owing to the attenuation of signal through the water column. When the image inversion process of our model is used to reverse the effects of attenuation and increase the signal gain appropriately, the noise in the images is also increased, such that the signal-to-noise of the corrected images is no better than the original image. This was particularly apparent in regions in images that were far from the lights and cameras, which displayed large amounts of noise. For image data in which the real signal is extremely attenuated (high altitude imaging or highly turbid water), it is likely that our correction methodology will not produce accurate colours or meaningful images.

Our correction methodology assumes that water attenuation and scattering coefficients are constant over the area of the dive. Our approach cannot account for changes that occur over space or time, for example in the situation that sediment is significantly disturbed during an imaging mission. We do not address the issues of contrast loss due to forward scattering of light through the water (see (Schechner and Karpel, 2004) for correction methodologies that address this issue). Although the reproduced colour accuracy is improved, the spatial resolution and spatial contrast of the images is not affected.

Our correction methodology only addresses situations in which the dominant light source is carried on-board the AUV, meaning our approach (in its current form) cannot be used when the dominant light source is from the sun or ambient light. Ambient light can potentially be added to the image formation model in Equation 13, however this would require explicit knowledge of the relative illumination power of the sun just below the water's surface to that of the artificial lights, which is difficult to measure and varies over the course of a day, is affected by weather conditions etc. This aspect is left as a topic for future work.

7 Conclusions and Future Work

This paper has presented an approach to reconstructing the true colour in underwater images by using an underwater image formation model (and its inverse) that accounts for water attenuation, scattering and artificial lighting spatial distribution. Parameters of the model are estimated from the image data itself, exploiting spatial information generated by using structure-from-motion and photogrammetry techniques, removing the requirement for colour calibration infrastructure to be used during image acquisition. Results were presented using two different AUV dives, demonstrating the ability of the method to recover accurate and consistent estimates of scene colours in different environments.

Our recent work has examined the use of hyperspectral sensors for measurements of spectral reflection on the seafloor (Bongiorno et al., 2013). In future work we plan to extend the image formation model derived in this paper to the hyperspectral domain, for use in colour correction and high-spectral resolution imaging. We will also extend our correction methodology to situations in which the underwater scene is illuminated by both artificial light and sun or ambient light.

Acknowledgments

This work is supported in part by Linkage Project LP0989998, funded by Thales Australia, the Australian Research Council (ARC) and the New South Wales State Government. The IMOS AUV facility is supported through the Department of Innovation, Industry, Science and Research (DIISR) National Collaborative Research Infrastructure Scheme. Special thanks to Christian Lees, Ariell Friedman and Daniel Bongiorno for their assistance during fieldwork.

References

- Beijbom, O., Edmunds, P., Kline, D., Mitchell, B., and Kriegman, D. (2012). Automated Annotation of Coral Reef Survey Images. In *IEEE Conference on Computer Vision (CVPR)*.
- Bewley, M., Douillard, B., Nourani-Vatani, N., Friedman, A., Pizarro, O., and Williams, S. (2012). Auto-

mated species detection: An experimental approach to help detection from sea-floor AUV images. In *Australasian Conference on Robotics and Automation*.

Bongiorno, D., Bryson, M., and Williams, S. (2013). Dynamic Spectral-Based Underwater Colour Correction. In *IEEE/MTS Oceans Conference*.

Bryson, M., Johnson-Roberson, M., Pizarro, O., and Williams, S. (2012). Colour-Consistent Structure-from-Motion Models using Underwater Imagery. In *Robotics: Science and Systems*.

Bryson, M., Johnson-Roberson, M., Pizarro, O., and Williams, S. (2013). Automated Registration for Multi-year Robotic Surveys of Marine Benthic Habitats. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

Buchsbaum, G. (1980). A Spatial Processor Model for Object Color Perception. *Journal of Franklin Institute*, 310(1):1–26.

Campos, R., Garcia, R., Alliez, P., and Yvinec, M. (2015). A surface reconstruction method for in-detail underwater 3D optical mapping. *The International Journal of Robotics Research*, 34(1):64–89.

Campos, R., Garcia, R., and Nicosevici, T. (2011). Surface reconstruction methods for the recovery of 3D models from underwater sites. In *IEEE OCEANS Conference*.

Chiang, J. and Chen, Y. (2012). Underwater image enhancement by wavelength compensation and dehazing. *IEEE Trans. on Image Processing*, 21(4):1756–1769.

Clarke, M., Tolimieri, N., and Singh, H. (2009). Using the SeaBED AUV to assess populations of groundfish in untrawlable areas. *The Future Fisheries Science in North America. Fish and Fisheries Series*, 1:357–372.

Davis, T. A. (2006). Direct Methods for Sparse Linear Systems. In *SIAM, Fundamentals of Algorithms*.

Eustice, R., Pizarro, O., and Singh, H. (2008). Visually Augmented Navigation for Autonomous Underwater Vehicles. *IEEE Journal of Oceanic Engineering*, 33(2):103–122.

- Finlayson, G., Hordley, S., and Hubel, P. (1998). Recovering Device Sensitivities with Quadratic Programming. In *The Sixth Color Imaging Conference: Color Science, Systems and Applications*, pages 90–95.
- Gracias, N., Ridao, P., Garcia, R., Escartin, J., L'Hour, M., Cibecchini, F., Campos, R., Carreras, M., Ribas, D., Palomeras, N., Magi, L., Palomer, A., Nicosevici, T., Prados, R., Hegedus, R., Neumann, L., de Filippo, F., and Mallios, A. (2013). Using a lightweight AUV to survey the site of the 17th Century ship 'La Lune'. In *IEEE OCEANS Conference*.
- Grasmueck, M., Eberli, G. P., Viggiano, D. A., Correa, T., Rathwell, G., and Luo, J. (2006). Autonomous underwater vehicle (auv) mapping reveals coral mound distribution, morphology, and oceanography in deep water of the straits of florida. *Geophys. Res. Lett.*, 33:6.
- Grossberg, M. D. and Nayar, S. K. (2004). Modelling the Space of Camera Response Functions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(10):1272–1282.
- Hochberg, E., Apprill, A., Atkinson, M., and Bidigare, R. (2006). Bio-optical modelling of photosynthetic pigments in corals. *Coral Reefs*, 25(1):99–109.
- Hochberg, E. and Atkinson, M. (2000). Spectral discrimination of coral reef benthic communities. *Coral Reefs*, 19(2):164–171.
- Jaffe, J. S. (1990). Computer Modeling and the Design of Optimal Underwater Imaging Systems. *IEEE Journal of Oceanic Engineering*, 15(2):101–111.
- Johnson-Roberson, M., Bryson, M., Douillard, B., Pizarro, O., and Williams, S. (2013). Out-of-core Efficient Blending for Underwater Georeferenced Textured 3D Maps. In *COM.Geo Conference, San Jose, USA, July*.
- Johnson-Roberson, M., Pizarro, O., Williams, S., and Mahon, I. (2010). Generation and Visualization of Large-scale Three-dimensional Reconstructions from Underwater Robotic Surveys. *Journal of Field Robotics*, 27(1):21–51.
- Kaeli, J., Singh, H., Murphy, C., and Kunz, C. (2011). Improving Color Correction for Underwater Image Surveys. In *IEEE/MTS Oceans Conference*.

- Kim, S. and Pollefeys, M. (2008). Robust Radiometric Calibration and Vignetting Correction. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 30(4):562–576.
- Lewis, G., Jordan, D., and Roberts, P. (1999). Backscattering target detection in a turbid medium by polarization discrimination. *Applied Optics*, 38(18):3937–3944.
- Lowe, D. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110.
- Mahon, I., Williams, S., Pizarro, O., and Johnson-Roberson, M. (2008). Efficient View-based SLAM using Visual Loop Closures. *IEEE Transactions on Robotics*, 24(5):1002–1014.
- Marquardt, D. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics*, 11(2):431–441.
- Mobley, C. (1994). Light and water: Radiative transfer in natural waters. In *vol. 592, Academic Press*.
- Pedrotti, F. and Pedrotti, L. (1993). Introduction to Optics. In *Prentice Hall, ISBN: 0135015456*.
- Schechner, Y. and Karpel, N. (2004). Clear Underwater Vision. In *IEEE International Conference on Computer Vision and Pattern Recognition*.
- Schettini, R. and Corchs, S. (2010). Underwater image processing: State of the art of restoration and image enhancement methods. *Journal on Advances in Signal Processing*, doi:10.1155/2010/746052.
- Sedlazeck, A., Koser, K., and Koch, R. (2009). 3D reconstruction based on underwater video from ROV Kiel 6000 considering underwater imaging conditions. In *IEEE OCEANS Conference*.
- Smale, D., Kendrick, G., Harvey, E., Langlois, T., Hovey, R., Niel, K. V., Waddington, K., Bellchambers, L., Pember, M., Babcock, R., Vanderklift, M., Thomson, D., Jakuba, M., Pizarro, O., and Williams, S. (2012). Regional-scale benthic monitoring for ecosystem-based fisheries management (EBFM) using an autonomous underwater vehicle (AUV). *ICES Journal of Marine Science*, 69(6):1108–1118.
- Soriano, M., Marcos, S., Saloma, C., Quibilan, M., and Alino, P. (2001). Image classification of coral reef components from underwater color video. In *Oceans Conference*.

- Tarel, J. and Hautiere, N. (2009). Fast visibility restoration from a single color or gray level image. In *IEEE 12th International Conference on Computer Vision*.
- Torres-Mendez, L. A. and Dudek, G. (2005). A Statistical Learning-Based Method for Color Correction of Underwater Images. In *Research on Computer Science Vol. 17, Advances in Artificial Intelligence Theory*.
- Vasilescu, I., Detweiler, C., and Rus, D. (2010). Color-Accurate Underwater Imaging using Perceptual Adaptive Illumination. In *Robotics: Science and Systems*.
- Williams, S., Pizarro, O., Jakuba, M., Johnson, C., Barrett, N., Babcock, R., Kendrick, G., Steinberg, P., Heyward, A., Doherty, P., Mahon, I., Johnson-Roberson, M., Steinberg, D., and Friedman, A. (2012). Monitoring of Benthic Reference Sites: Using an Autonomous Underwater Vehicle. *IEEE Robotics and Automation Magazine*, 19(1):73–84.
- Yamashita, A., Fujii, M., and Kaneko, T. (2007). Color Registration of Underwater Images for Underwater Sensing with Consideration of Light Attenuation. In *IEEE International Conference on Robotics and Automation*.
- Yoerger, D., Jakuba, M., Bradley, A., and Bingham, B. (2007). Techniques for deep sea near bottom survey using an autonomous underwater vehicle. *International Journal of Robotics Research*, 26:41–54.

	Iver AUV	Sirius AUV
Size	2.0x1.5x1.5m	2.0x1.5x1.5m
Mass	45kg	200kg
Depth Rating	100m	800m
Camera Type	Prosilica 12B CCD, 1360x1024pix	Prosilica 12B CCD, 1360x1024pix
Frame Rate	2Hz	2Hz
Exposure	manual, constant	manual, constant
Artificial Lights	Dual Strobes (LED Array)	Dual Strobes (Xenon Flash)
Typical Altitude	2-4m	1-3m
Other Sensors	Depth, DVL	Depth, DVL, USBL

Table 1: Specifications of AUV Platforms and Sensors used in the experimental setup.