An underwater lighting and turbidity image repository for analysing the performance of image based non-destructive techniques

Michael O'Byrne[1,2], Franck Schoefs[3], Vikram Pakrashi[1,2]*, Bidisha Ghosh[4]
[1] Dynamical Systems and Risk Laboratory, School of Mechanical and Materials Engineering, University College Dublin, Belfield, Dublin 4, Ireland
[2] Marine Renewable Energy Ireland (MaREI) Ireland, University College Dublin
[3] Université Bretagne-Loire, Université de Nantes, Research Institute of Civil Engineering and Mechanics (GeM)/Sea and Littoral Research Institute (IUML), CNRS UMR 6183/FR 3473, Nantes, France; IXEAD/CAPACITES Society, Nantes, France
[4] Department of Civil, Structural and Environmental Engineering, Trinity College Dublin, Ireland
Email: michael.obyrne@ucc.ie, franck.schoefs@univ-nantes.fr, *vikram.pakrashi@ucd.ie (corresponding author), bghosh@tcd.ie

Abstract

Image processing based methods, capable of detecting and quantifying cracks, surface defects or recovering 3D shape information are increasingly being recognized as a valuable tool for inspecting underwater structures. It is of great practical importance for inspectors to know the effectiveness of such techniques when applied in conditions. This paper considers an underwater environment characterised by poor visibility chiefly governed by the lighting and turbidity levels, along with a range of geometry and damage conditions of calibrated specimens. The paper addresses the relationship between underwater visibility and the performance of image based methods through the development and calibration of a first open-source Underwater Lighting and Turbidity Image Repository (ULTIR). ULTIR contains a large collection of images of submerged specimens that have been photographed under controlled lighting and turbidity levels featuring various forms of geometry and damage. ULTIR aims to facilitate inspectors when rationalising the use of image processing methods as part of an underwater inspection campaign and to enable researchers to efficiently evaluate the performance of image based methods under realistic operating conditions. Stakeholders in underwater infrastructure can benefit through this first large, standardized, well-annotated, and freely-available database of images and associated metadata.

Keywords: non-destructive testing; marine and offshore engineering; image analysis, underwater visibility; calibration; defects; databases; maintenance & inspection, standardisation

1.0 Introduction

Offshore and marine structures have developed rapidly in recent decades. Much of this has been propelled by our growing appetite to harness our ocean wealth and through continued

advances in construction materials and methods. Such advances have enabled structures to be installed at an increasing rate numbers and in deeper waters. The harsh and unrelenting marine environment means that these structures are particularly susceptible to aesthetic, functional or structural degradation, which over time, typically leads to a loss of serviceability at either a component or global level. Owners/managers must therefore inspect structures to ensure that they are safe, fit-for-service, and so that they can make more informed decisions when allocating resources towards the correction of deficiencies. This latter aspect has attracted a growing interest in recent times as the importance of life cycle optimisation and related financial benefits continue to be recognized, especially in relation to marine structures (Schoefs et al, 2012a).

Assessing the submerged part of marine structures introduces new challenges for inspectors. Many damage diagnostic tools that can be used on dry land cannot be readily adapted for underwater deployment. Additionally, only a limited amount of time can be spent underwater, especially when the inspection is being carried out by a diver rather than an ROV (Remotely Operated Vehicle). This puts an emphasis on adopting expeditious data collection practices that harness the full potential of available tools. A comprehensive summary of the challenges facing underwater inspections is provided by Busby (1979) and Ramos (1992), which still remain relevant despite being published decades ago.

The most common form of underwater inspection is a visual inspection carried out by trained divers (Anderson, 1987). Visual inspections are affected by the ability of inspectors to observe and objectively record details of defects, and are prone to considerations such as boredom, lapses in concentration, subjectivity, and fatigue, all of which contribute to greater variability and reduced accuracy (Agin, 1980, Komorowski and Forsyth, 2000, Estes and Frangopol, 2003, Dirksen et al., 2011).

An increasingly popular way of yielding more data from inspections, and in the process, offsetting some of the inherent shortcomings of conventional visual inspections, relies on using cameras in conjunction with sophisticated image analysis algorithms to make visual data a part of quantitative assessment. The quantitative nature of the data obtained from image analysis naturally lends itself to numerous applications and is helpful for developing new damage models, or strengthening existing ones, which are used to forecast the rate of propagation of damage as the structure continues to operate. Unfortunately, underwater imaging is hampered by poor visibility which diminishes the ability of the camera and subsequent image processing algorithms to effectively identify and quantify instances of damage. The extensive effort and expense associated with underwater inspections mean that inspectors can ill-afford to rely on unproven methods (Schoefs et al., 2012b). It is therefore important that inspectors can form a relationship between the on-site visibility conditions and the performance of image-based techniques.

The degree of underwater visibility is largely dependent on the lighting and turbidity levels. Turbidity is defined as the cloudiness in a liquid caused by the presence of suspended solids that scatter and absorb light and therefore reduce visibility. To date, there has been little work carried out on characterising the performance of image based methods applied in an underwater setting. Bianco et al. (2013) studied the effects of lighting and total suspended solids (TSS) on a single stereo vision technique and O'Byrne et al. (2013a) investigated the effects of turbidity and lighting on a surface damage detection technique. Total suspended solids, which is the mass of suspended particles per volume of water, is related to turbidity in the sense that more suspended particles lead to increased cloudiness in the water. However, turbidity also takes into account relevant properties such as particle shape, refractive index and the colour of suspended solids. These properties vary from site to site meaning that TSS cannot be universally used as a metric for measuring water clarity, unlike turbidity. There is

thus significant scope in drawing a link between the on-site conditions and the performance of image based methods in a systematic and meaningful manner.

This research presents a repository driven approach to comprehensively map the effect of turbidity and lighting on the performance of image based techniques. Such a repository is timely given the increasing need for underwater inspections worldwide and the growing popularity of image based methods as part of the inspection framework. The repository is termed Underwater Lighting and Turbidity Image Repository (ULTIR). It may be accessed through an intuitive web-interface (available at: http://www.ultir.net). It is populated with images featuring various damage forms and material types that have been photographed under controlled lighting and turbidity levels, reflective of real-world operating conditions. The purpose of ULTIR is to provide inspectors and researchers with insight into the relationship between underwater visibility and the performance of image processing based techniques. Addressing this knowledge gap has substantial practical implications - it enables inspectors to assess the viability of adopting image processing based approaches prior to an inspection and helps them to identify the limits at which image based methods begin to produce unacceptably poor results. Additionally, the information gleaned from applying algorithms to images in ULTIR can be used during actual inspections to create conditions that are most conducive to good detection such as the provision of suitable lighting - something that is easily attainable. Finally, although ULTIR is primarily aimed at image-based damage assessment algorithms, this resource may also find use for testing other image algorithms such as image de-noising and contrast enhancement.

The following section details the design and content of ULTIR. Section 3 demonstrates the application of image based techniques to imagery from the repository and develops the technique evaluation facet by comparing several techniques, including general techniques as well established techniques in the domain of NDT. Section 4 concludes the paper.

2.0 Design and Content

ULTIR consists of images generated from experiments conducted in a controlled underwater setting. This section describes the contents of the repository, the experimental set-up used to generate the imagery, and the web-interface where the imagery can be accessed. A schematic of the design and contents of the repository is shown in Figure 1.

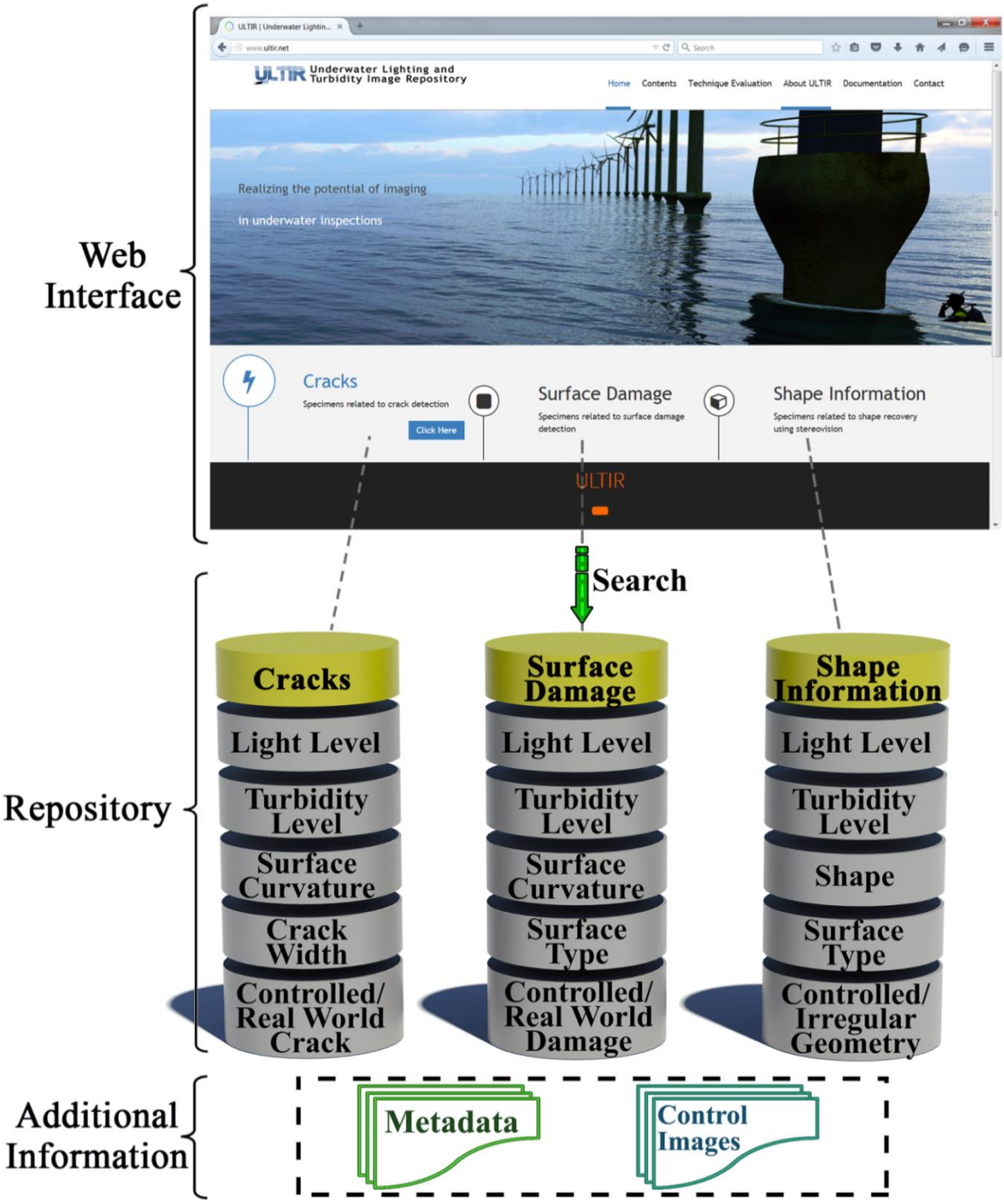

Figure 1. Web-interface and contents of the repository.

This figure presents the web-page that users will first encounter. From here they can find links to the imagery and metadata in the repository. The imagery is organized into three damage categories related to: cracks, surface damage, and 3D shape information. There is metadata associated with each image. It includes EXIF (Exchangeable Image File Format) information that reveals camera settings such as the shutter speed and ISO, along with concise contextual information about the specimens under consideration such as their size, shape, material, and damage description, where appropriate.

2.1 Contents of the Repository

ULTIR contains a large and diverse array of images covering a wide range of conditions and situations encountered underwater. This diversity is necessary for proper and comprehensive validation of algorithms. Table 1 provides a breakdown of the contents.

Table 1 Breakdown of the contents of the repository.

Section	No. of Specimens	Level of Control	Surface Type	Shape/ Curvature	Light Levels	Turbidity Levels	No. of images
Cracks	9	7 controlled, 2 real cracks	8 concrete, 1 textured concrete	Surface curvature: 4 flat, 5 curved	3 levels: 100 lux, 1000 lux, 10000 lux	3 levels: 0 NTU, 6 NTU, 12 NTU	81
Surface damage	10	9 controlled, 1 real damage	4 concrete, 3 textured concrete, 3 metallic	Surface curvature: 4 flat, 3 cylindrical,	3 levels: 100 lux, 1000 lux, 10000 lux	3 levels: 0 NTU, 6 NTU, 12 NTU	90
3D shapes	12	9 controlled, 3 irregular shapes	4 concrete, 4 metallic, 3 plastic, 1 rubber	3 cubes, 3 cylinders, 3 spheres, 3 irregular	3 levels: 100 lux, 1000 lux, 10000 lux	3 levels: 0 NTU, 6 NTU, 12 NTU	(108 x 8) 864

The images are defined by several attributes - most notably the turbidity and lighting conditions under which the image was captured and the type of damage under consideration - but also by other attributes such as the object curvature and surface type. The significance of each of these attributes is discussed in greater detail later in this section.

Although there are a similar amount of specimens used in each section, it may be noted from Table 1 that there are a significantly more images in the 3D shape section of the repository than in the crack and surface damage sections. This is due to the fact that recovery of 3D shape requires multiple photographs to be captured from different perspectives. In total, four stereo pairs (each consisting of two photographs from the left camera and right camera) were captured from various sides around the specimens to enable full 3D shape recovery, thereby resulting in eight images per specimen.

2.2 Controlled and Partially Controlled

The specimens are classified as either controlled or partially controlled. For the controlled specimens, the 'damage' is artificially created such that its location and true dimensions are precisely known. Introducing controlled damage in such a manner is an established way of accurately evaluating the performance of damage detection techniques as it reduces measurement uncertainties that generally arise when assessing real-world instances of damage where uncertainties and vagaries are usually present (Hearn and Testa, 1991). The partially controlled specimens feature real world instances of damage or, in the case of the 3D shape section of ULTIR, irregular shaped objects are used. For the partially controlled specimens, precise knowledge of the size of the damage is not known beforehand and must be visually identified by a human observer from the images. The visually segmented images are assumed to show the true extent of damage. Approximately 80% of the images in ULTIR represent controlled specimens while the remainder are partially controlled specimens. The

controlled specimens have the primary purpose of algorithm validation, while the partially controlled are intended more so for testing.

2.3 Damage Type

The repository is partitioned according to the type of damage. There are three general damage forms considered, namely; cracks, surface damage, and damages quantified by 3D shape. The background and nature of each of these damage forms is discussed below:

2.3.1 Cracks

Cracks provide an indication of the structural degradation and are an important factor when diagnosing the condition of concrete and steel structures. The measurement of cracks on offshore structures using various non-destructive tools has been studied in the past, especially in relation to steel structures (Rouhan and Schoefs, 2003; Barnouin et al., 1993; Rudlin, 1996). However, cracks on concrete structures are more difficult to measure as the shape tends to be more irregular which is why this repository focuses mainly on concrete specimens. Moreover, concrete marine structures are especially susceptible to cracks due to the high dynamic loading and the presence of corrosive salts that are absorbed into the concrete and corrode the reinforcing steel, causing volume expansion and cracking.

Traditional visual monitoring methods typically require an inspector to map, count, quantitatively measure, and photograph the observed cracks. This tedious process can be significantly improved through image processing based methods which can automatically count, classify and quantify the length and width of cracks, thereby enhancing inspections, and in turn, leading to significant monetary savings or leading to more efficient inspection cycles. Although it is also sometimes possible to estimate the depth of cracks or fissures using image based techniques, this aspect is currently beyond the scope of this repository.

Image based crack detection algorithms, such as (Iyer and Sinha, 2006, Mei et al., 2004, Nishikawa et al., 2012; Lee et al., 2013; Omondi et al., 2016; Ebrahimkhanlou et al., 2016;), work by identifying features of cracks such as their narrow shape and their lower brightness

in comparison to the surroundings. For the controlled part of the repository, cracks were simulated by offsetting two halves of split concrete specimens by a known distance. The controlled cracks range in width from 1 mm to 5 mm. These simulated cracks share all the features of real cracks such as the fine structure and lower brightness relative to the surroundings, as shown in Figure 5.2. The images shown were captured under medium lighting (1,000 lux) and in clear water (0 Nephelometric Turbidity Units (NTU)) conditions. ULTIR contains two cases of real-world cracks on different surfaces (Figure 5.2(h-i)).

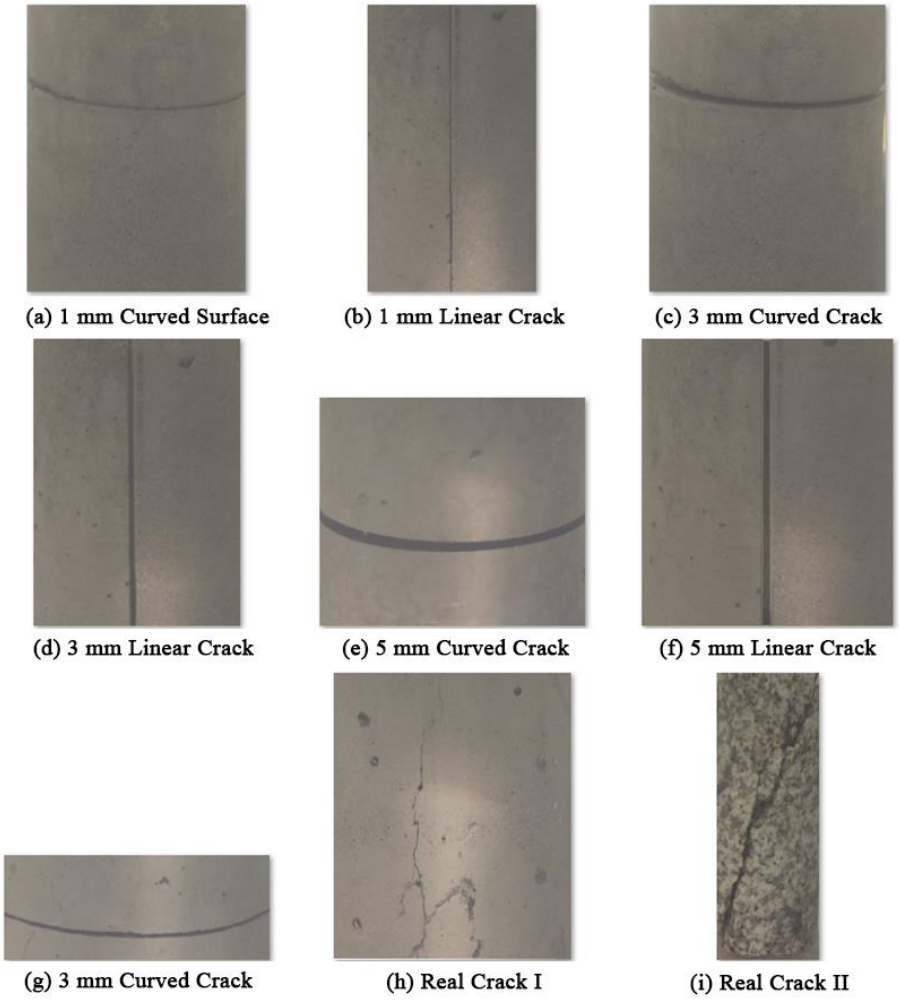

Figure 2 Specimens in the crack section of the repository. These illustrated images were captured under medium light (1,000 lux) and low turbidity (0 NTU) conditions.

2.3.2 Surface Damage

The aim of the image based 2D surface damage algorithms is to locate and quantify the area occupied by visible damage (typically larger than 10^{-6} m^2) on the surface of infrastructural

elements with minimal human supervision. Marine structures are affected by a wide range of visible damage forms (Dirksen et al., 2014; Valença et al., 2014) such as corrosion, scour, erosion, leaching, spalling, impact damage etc. The contents of the repository reflect this variety by including specimens with different surface textures, as shown in Figure 3.

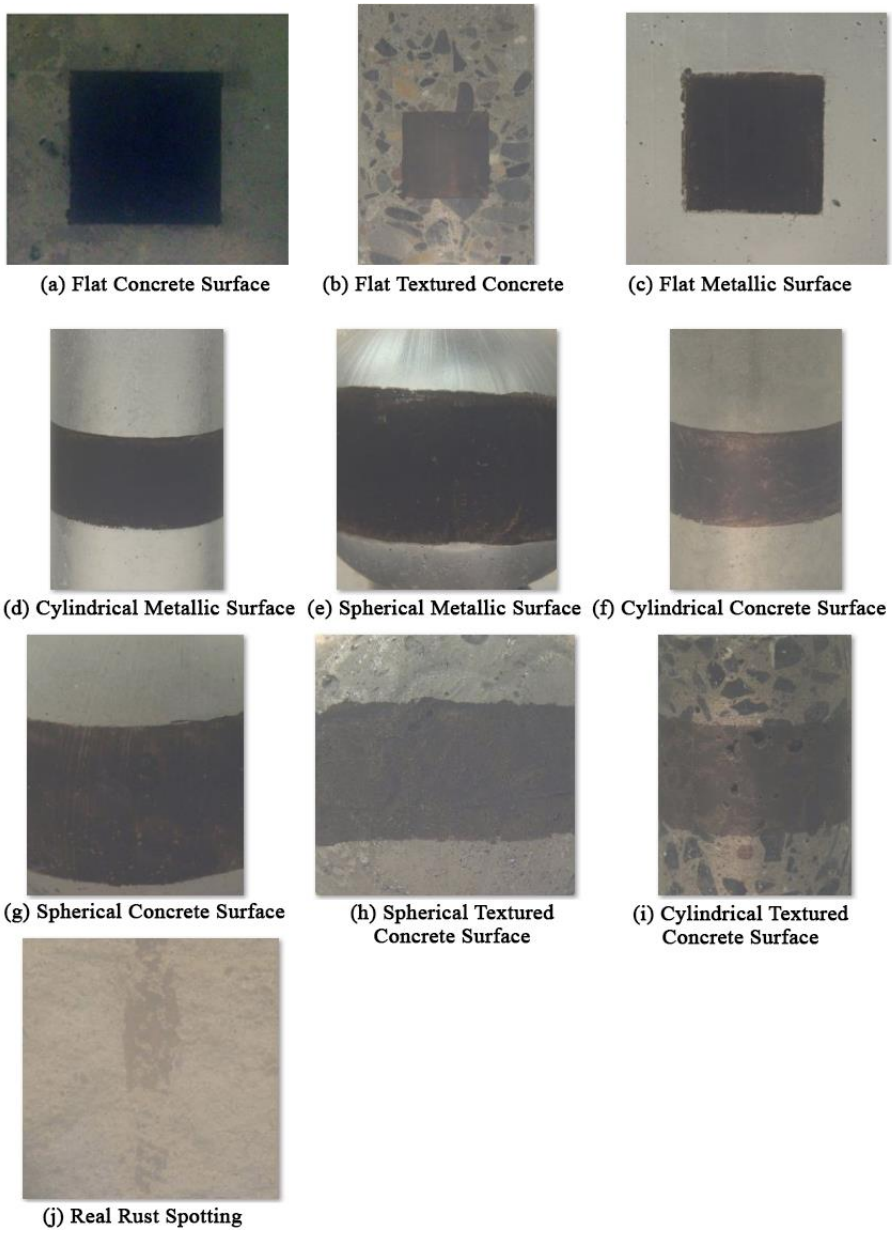

Figure 3 Specimens in the surface damage section of the repository. These images were captured under medium light (1,000 lux) and low turbidity (0 NTU) conditions.

These damages come in an array of shapes, sizes and pose varying levels of significance to the health of a structure. While some damage forms such as pitting corrosion are generally

detected using colour information, other damages such as spalling and erosion may be more differentiable based on their textural properties in comparison to the undamaged surface. Image based damage detection algorithms have been developed to segment based on colour information (Yazid et al., 2011, O'Byrne et al., 2014b) and textural information (O'Byrne et al., 2013b).

For the controlled part of the repository, damage is simulated by applying a typically rust-coloured paint to standard geometric shapes such as squares and rectangles of known dimensions on different surfaces and surface curvatures. The repository also contains a real-world case of rust spotting (Figure 3(j)).

2.3.3 Shape Information

Recovering shape information is a challenging but useful task, and has wide applicability in many areas of Structural Health Monitoring (SHM). For underwater inspections, one such example concerns the tracking of marine growth thickness on offshore structures. Marine growth is undesirable as it increases the hydrodynamic forces acting on the host structure (Boukinda Mbadinga et al., 2007). There are a number of practical image based approaches capable of recovering 3D shape. These are structured lighting, Structure from Motion (SfM), and stereo photography. Structured lighting techniques utilize a light projector and a camera to project a light pattern onto an object and capture how it interacts with the shape (Bruno et al., 2011). Given the reliance on a light source to encode depth information, the success of this approach is particularly susceptible to absorption and scattering.

Structure from Motion approaches, such as (Li et al., 1997), can obtain 3D shape information from a single camera moving in a static scene. Features in the scene are tracked from one image to the next and their trajectories over time are then used to reconstruct their 3D positions and the camera's motion. However, SfM algorithms require a scene to remain relatively rigid which is not always possible in an underwater setting as deforming surfaces,

floating particulate, and illumination changes are all sources of non-rigidity. Additionally, SfM has been shown to be less reliable than stereo vision based methods when applied underwater (O'Byrne et al. 2014c).

Stereo vision is a popular way of obtaining shape information, and has previously been applied in an underwater setting by Brandou et al. (2007). A stereo system consists of two synchronized cameras viewing the scene from two slightly different vantage points. The captured images from both cameras are collectively known as a stereo pair. By examining the relative positions of objects in each image, 3D information can be extracted. Stereo systems are capable of providing a fully scaled metric scene reconstruction once the system is calibrated, using either standard techniques (Zhang, 2000) or by auto-calibration (Faugeras et al., 1992). A fully scaled metric reconstruction is an accurate representation of the scene; true angles are preserved and object dimensions are in real world units. ULTIR is populated by stereo pairs of numerous specimens. The controlled specimens are standard geometric shapes in the form of spheres, cubes and cylinders, as shown in Figure 4. The controlled specimens were chosen as these primitive shapes are the building blocks for more complex shapes. Therefore, it is important to get a fundamental understanding of how stereo-matching algorithms handle the various curvatures. The uncontrolled specimens are irregular shapes with more intricate depth variations which is more reflective of what would be encountered during real-world inspections.

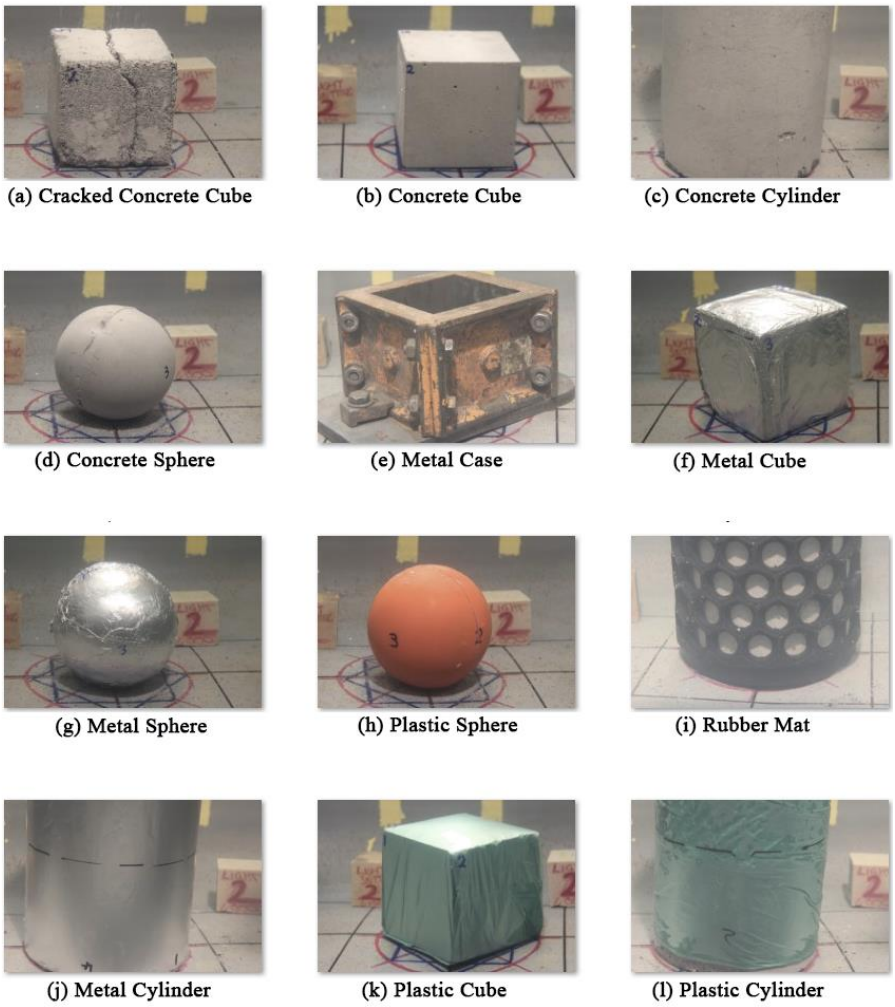

Figure 4 Specimens in the surface damage section of the repository. These images were captured under medium light (1,000 lux) and low turbidity (0 NTU) conditions.

2.3.4 Turbidity and lighting

Image quality is assumed to be chiefly affected by luminosity, sharpness (focus accuracy), contrast and noise. These quality factors are directly related to the on-site operating conditions, for which lighting and turbidity are the most influential (Mahiddine et al., 2012). Underwater imaging is severely hampered by turbidity, which results in reduced contrast, loss of details, and colour alteration. Turbidity can be caused by organic particles, such as decomposed plant and animal matter, and algae; or by inorganic particles such as silt and clay. In rivers and lakes, the level of turbidity can fluctuate due to a number of factors such as heavy rains and urban run-off, landslides and bank erosion, algae blooms, interference during the inspection process (e.g. disturbing sediment on the river bed), and human activities such

as construction. Given these factors, significant variations in turbidity can be expected. For example, the measured turbidity in the lower Waitaki river, New Zealand, ranges from 1.2 NTU (Nephelometric Turbidity Units) to 23 NTU (Graham, 1990). With this in mind, inspectors may wish to schedule inspections for periods when these factors are expected to cause minimum increases in turbidity levels. In the open ocean, turbidity is affected mostly by seasonal phytoplankton blooms, however it is generally low.

For ULTIR, three levels for turbidity were chosen: 0 NTU, 6 NTU, and 12 NTU. Clear water has a turbidity of 0 NTU, water that is visibly cloudy has a turbidity of 6 NTU, while water that is murky has a turbidity of 25 NTU. A cut-off point of 12 NTU was chosen as it becomes increasingly difficult to interpret and extract useful information from images beyond this point. In waters above 12 NTU, image based methods become an increasingly infeasible option as quantitative inspection tool. Additionally, the turbidity of many rivers and water bodies' lies within the 0 – 12 NTU range, and thus, studying this range is of high practical relevance. For example, Grand River, Michigan is reported to have a turbidity range of 2 – 9 NTU, while Spring Lake in New Jersey has a turbidity range of 0.1 – 4 NTU (GVSU, 2013). The lower limit of 0 NTU was chosen as this represents the best case scenario and it shows the level of performance can be achieved in the absence of any turbidity, while 6 NTU was a natural choice as the intermediate turbidity level as it is the midpoint of the 0 – 12 NTU range.

Turbidity can be measured on site using a digital nephelometer or by using a Secchi disk which is a black and white disk lowered into the water until it is no longer visible. The depth of the disk is a measure of the transparency of the water which is inversely related to the turbidity. This has the advantage of being a quick, inexpensive and simple approach for measuring turbidity. Additionally, many water bodies, such as the River Lee in Ireland, are

increasingly being monitored by arrays of sensor networks, which provide near real-time data on water quality parameters including turbidity (Lawlor et al., 2012).

While lowering the on-site turbidity levels is generally not possible, the effects of high turbidity on image quality can be partially offset by moving the camera closer to the subject. In high-turbid/murky waters, e.g. above 100 NTU, an object may disappear if it is separated by just a few centimetres from the observer, while in clearer waters, such as in the open Ocean where the turbidity is often close to 0 NTU, an object can be tens of meters away from an observer and still remain visible.

One way of explaining this phenomenon is by looking at the interaction between light rays and suspended solids in the water, which are the primary cause of turbidity (Fondriest, 2014). When photographing an object in turbid waters, i.e. water with a high concentration of suspended solids, light rays will strike some of the suspended particles between the object and the cameras and will become scattered. This leads to the object having a cloudy or fuzzy appearance in the eventual photographs. By moving the camera closer to the object, the light rays travel through less of the turbid medium, and as a consequence, the rays will collide with fewer suspended particles. The object will then have a more well-defined appearance in the photographs. With this in mind, the effects of turbidity can be mitigated by moving the camera closer to the object; however, this must be balanced by practical and logistical considerations. For instance, if the camera is too close to the object then the resulting imagery loses context and is only capable of capturing small areas of the scene. The imagery in ULTIR is generated with the cameras and specimens kept at a fixed distance of 80 cm apart. This distance is regarded to be a practical distance that underwater divers would find to be a reasonable compromise between capturing enough of the scene without sacrificing too much detail.

Lighting also plays a pivotal role for achieving good visibility. Ambient light may be sufficient for near-surface inspections; however, it is unlikely to be sufficient at greater depths at which point artificial light will become necessary. Three light levels were used: 100 lux, 1000 lux, 10000 lux. To put this in perspective, the approximate level of light, or illuminance, on a very dark overcast day is 100 lux, a moderately overcast day is 1000 lux, and full daylight (not direct sunlight) is 10,000 - 25,000 lux (Schlyter, 2009). The lux is the SI unit of illuminance that measures the intensity of light that strikes a surface, as perceived by the human eye. It is similar to irradiance, which has units of watts per square metre, but with the power at each wavelength weighted according to a standardised luminosity function that is based on a model of human visual brightness perception. The illuminance can be measured using a lux meter, which are readily available and portable. A specimen from each section of the repository is shown under varying lighting and turbidity levels in Figures 5 - 7.

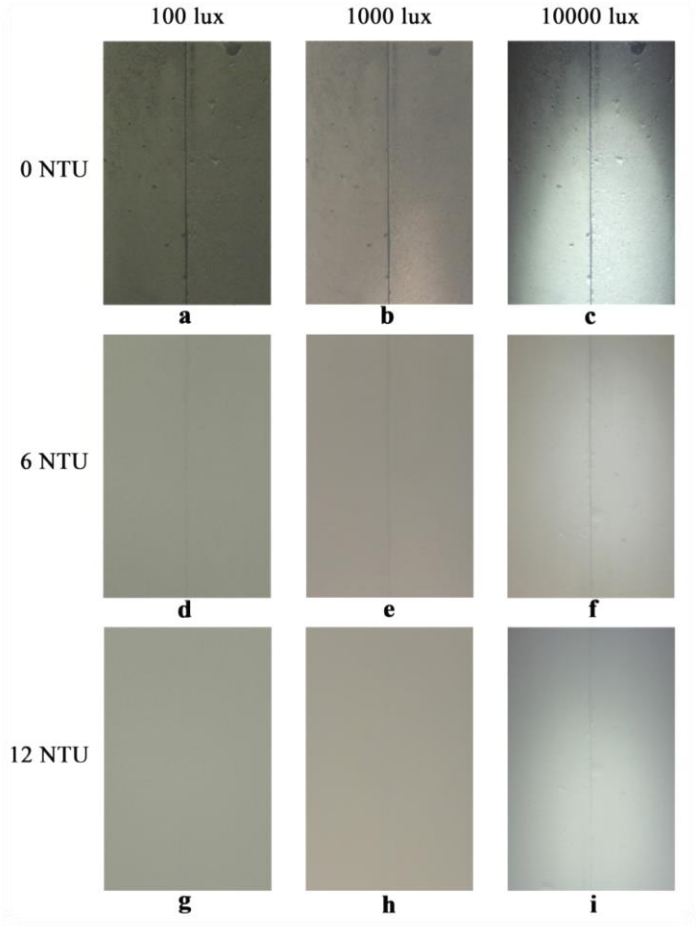

Figure 5 Controlled crack specimen shown under varying lighting and turbidity conditions. Columns: Low (100 lux), Medium (1000 lux), High (10000 lux) Light. Rows: Low (0 NTU), Medium (6 NTU), High (12 NTU) Turbidity.

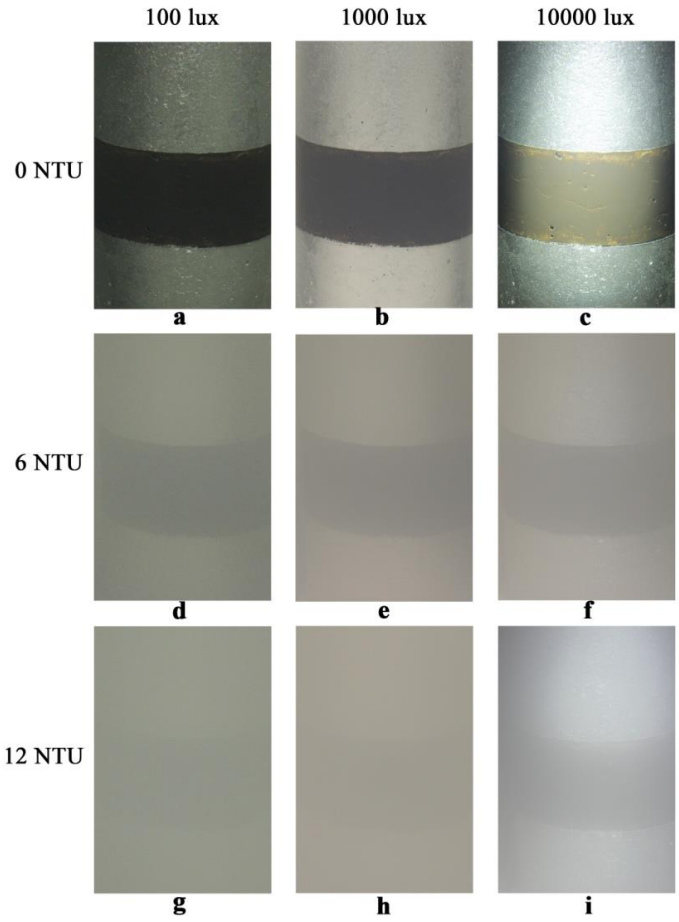

Figure 6 Controlled surface damage specimens under varying lighting and turbidity conditions. Columns: Low (100 lux), Medium (1000 lux), High (10000 lux) Light. Rows: Low (0 NTU), Medium (6 NTU), High (12 NTU) Turbidity.

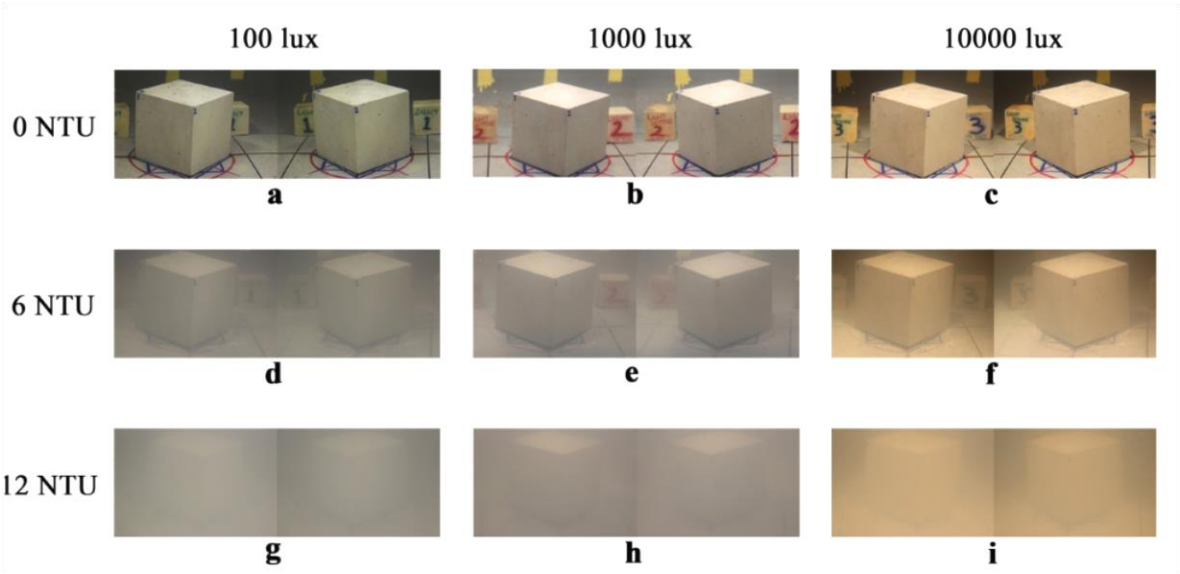

Figure 7 Stereo pairs featuring a concrete cube shown under varying lighting and turbidity levels. Columns: Low (100 lux), Medium (1000 lux), High (10000 lux) Light. Rows: Low (0 NTU), Medium (6 NTU), High (12 NTU) Turbidity.

It may be observed from Figures 5-7 that the featured specimens are easily distinguishable under some conditions (e.g. medium light with low turbidity), while in other cases, the appearance of the specimens are heavily degraded (e.g. low light with high turbidity). While it is clear that increasing turbidity has a worsening effect on image quality, the relationship between light intensity and image quality is more complicated. A dark, poorly lit scene is clearly unfavourable for the task of visual interpretation of damage; however, too much lighting can also create problems. With reference to the metallic specimen shown in Figure 6, it can be seen that lighting complexities, such as bright spots, become apparent at the highest lighting level. This obscures detail in affected regions of the image. Furthermore, high lighting in combination with high turbidity produces a lot of backscatter resulting in a hazy-glow effect, which also masks detail. This phenomenon is evident in Figure 6(i). The interplay between lighting and turbidity means that these must be considered jointly for the purpose of assessing the performance of image-based algorithms.

2.3.5 Surface type

The main construction materials for marine structures are concrete and metal. These materials have different photometric properties. Metallic surfaces often appear shiny to the eye due to a large amount of light being reflected in a specular fashion whilst concrete surfaces reflect very little specular light, instead reflecting light in a scattered fashion which produces a dull appearance (Dana et al., 1997). High specular reflections pose problems for image analysis algorithms as the shine masks details in the scene and create artefacts that could mislead algorithms. This is especially problematic when strong light sources are used.

The texture of a surface is another important property that has an effect on image analysis. Texture may be qualified by terms such as fine, coarse, smooth, rippled, molled, irregular, or lineated (Haralick et al., 1973). The influence of texture can be seen during the stereo-matching process in which corresponding points in two images must be matched. Matching points on smooth or uniform surfaces is more ambiguous than matching points on coarsely textured surfaces. Additionally, the surface texture may be a consideration when deciding on what type of surface damage detection algorithm to use. If damage is characterized more so by textural composition than by colour from the undamaged surface then it may be worthwhile segmenting based on texture. Finally, crack detection algorithms applied to surfaces with a rippled texture can produce a lot of false alarms. To cater for these issues, ULTIR contains metallic surfaces with high specular reflectivity as well as diffuse concrete surfaces with various textural finishes.

2.4 Experimental Set-up

The images in ULTIR were generated from experiments that were conducted in a water basin. These experiments were run in two phases. The set-up for the first phase entailed having a single underwater camera focused on damaged specimens. This phase produced the imagery for the crack and surface damage sections of the repository. The second phase employed on a dual-camera set-up as shown in Figure 8. This phase produced the stereo imagery for the 3D shape information section of the repository.

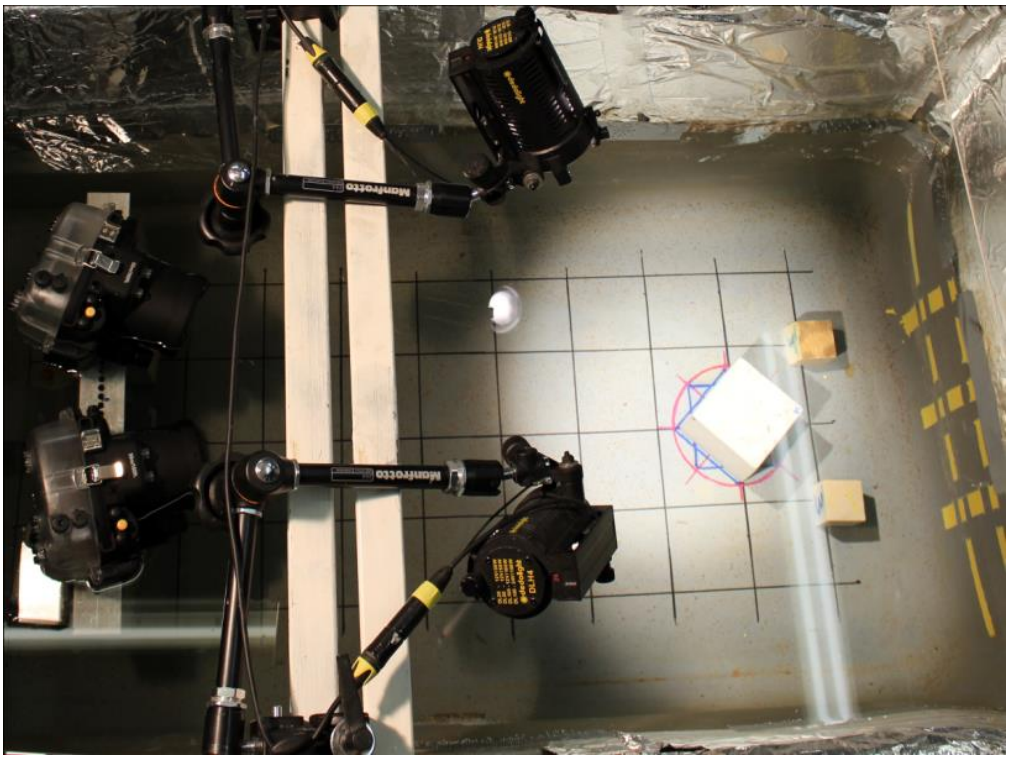

Figure 8. Plan view of experiment set-up.

2.4.1 Controlling the turbidity and light levels

Both phases of the experiment followed the same process in terms of controlling turbidity and illuminance levels. A flowchart of the test process is shown in Figure 9.

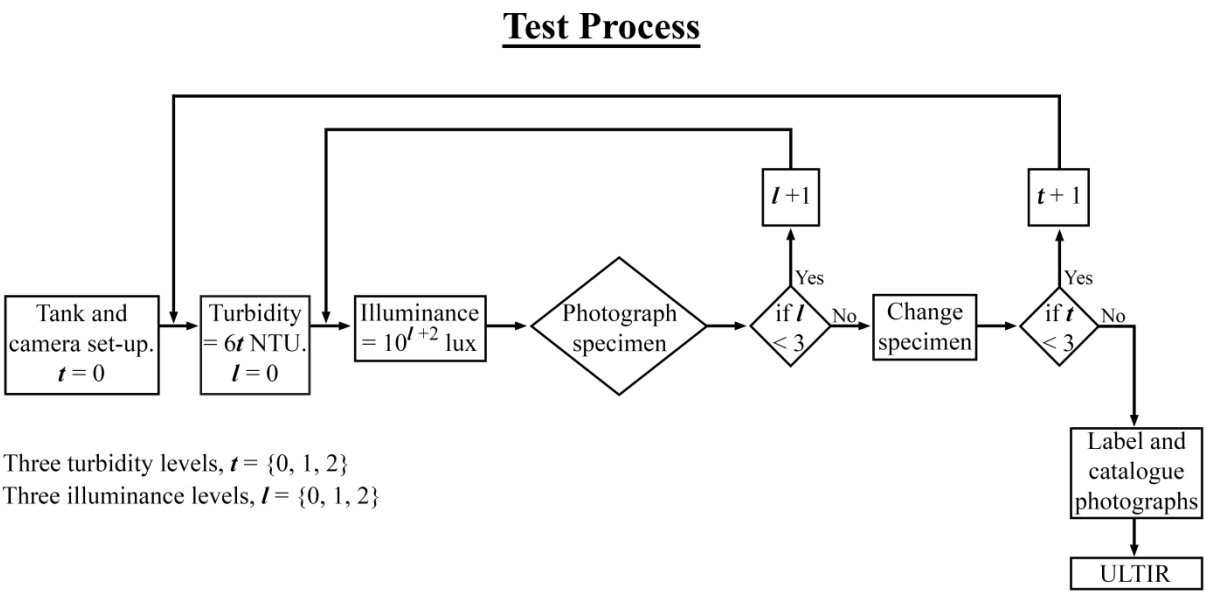

Figure 9. Flowchart of the test process

The turbidity was measured using a digital turbidimeter (model: Hach 2100P), which operates by measuring the loss of intensity of transmitted light through a water sample. Initially, the basin was filled with clear water. After all the photographs were captured at this turbidity level for each illiminance level, finely sieved kaolin was added in order to bring the turbidity up to 6 NTU, and later, to 12 NTU. Kaolin is a soft white clay consisting principally of the mineral kaolinite which goes into suspension when mixed with water. Regular stirring was carried out to ensure kaolin remained in suspension and was uniformly distributed.

The light levels were measured using a lux meter (model: AideTek LX1330B). Ambient lighting and a set of artificial lights were used to produce light levels of 100, 1000, and 10000 lux. Light was measured at the same position just above the specimen. While the illumination over the surface of a specimen will vary somewhat as a result of the slightly different distances from the light sources, the lux values span multiple orders of magnitude so slight variations within a given light level are negligible compared to differences between levels.

2.4.2 Camera parameters

The imagery was captured using two DSLR (Digital Single Lens Reflex) cameras, namely the Canon EOS 600D (Canon Rebel T3i), with standard kit lens (Canon EF-S 18–55mm lens f/3.5–5.6), which were enclosed in underwater housings. There are a number of camera settings which must be configured in order to get the desired quality of imagery. These settings include the shutter speed (exposure time), aperture, and ISO. The aperture, in conjunction with the focal length, determines the depth of field, while the shutter speed regulates the amount of motion blur. Together, the shutter speed and aperture control how much light arrives at the camera sensor. In low light conditions such as an underwater s, a slow shutter coupled with a large aperture is necessary to ensure sufficient light exposure.

The other important setting is the ISO which dictates the sensitivity of a camera's sensor to the available light. The lower the ISO number, the less sensitive the camera sensor is to incoming light, while a higher ISO number increases the sensitivity. Higher ISO settings are generally used in darker situations to allow for faster shutter speeds, however, raising the ISO means a similar decrease in quality, with an increase in noise. While early digital cameras had objectionable levels of noise at ISOs as low as 800, most modern DSLRs produce good quality images at ISOs up to 1600 and above. A balance must be struck between all of these camera settings. The shutter speed was chosen as the critical parameter as diver-held cameras employed in underwater inspections are prone to shaking which leads to unacceptable levels of motion blur if long shutter speeds are being used. For this reason, shutter priority mode was used. This is a semi-automatic shooting mode that allows the user to specify the shutter speed while the camera automatically decides the best aperture and sometimes ISO sensitivity to achieve the correct exposure. For ULTIR, the shutter speed was set to remain faster than 1/15 seconds.

The second phase of the experiment that involves collecting stereo pairs for the purpose of 3D shape recovery uses the two cameras at once. This phase required both cameras to operate

simultaneously and to each have the same camera parameter configuration. The cameras were synchronised using a remote control that was linked to both cameras. Once this remote was triggered, the cameras began photographing and continued to capture photographs at a pre-set time interval. The cameras were configured to save images in RAW and JPEG format with pixel dimensions of 5184 x 3456. The JPEG format is significantly smaller in terms of file size compared to the RAW format with little perceptible loss in image quality. For this reason, the JPEG format is used to disseminate the images via the web-interface for ease of transmission.

2.5 Web-Interface

Users of the repository will first encounter the web-interface. The design of the web-interface is focused on facilitating navigation through the repository. An extensive overview of the repository and a user guide is available at the ULTIR website. Some of the key web pages are shown in Figure 10.

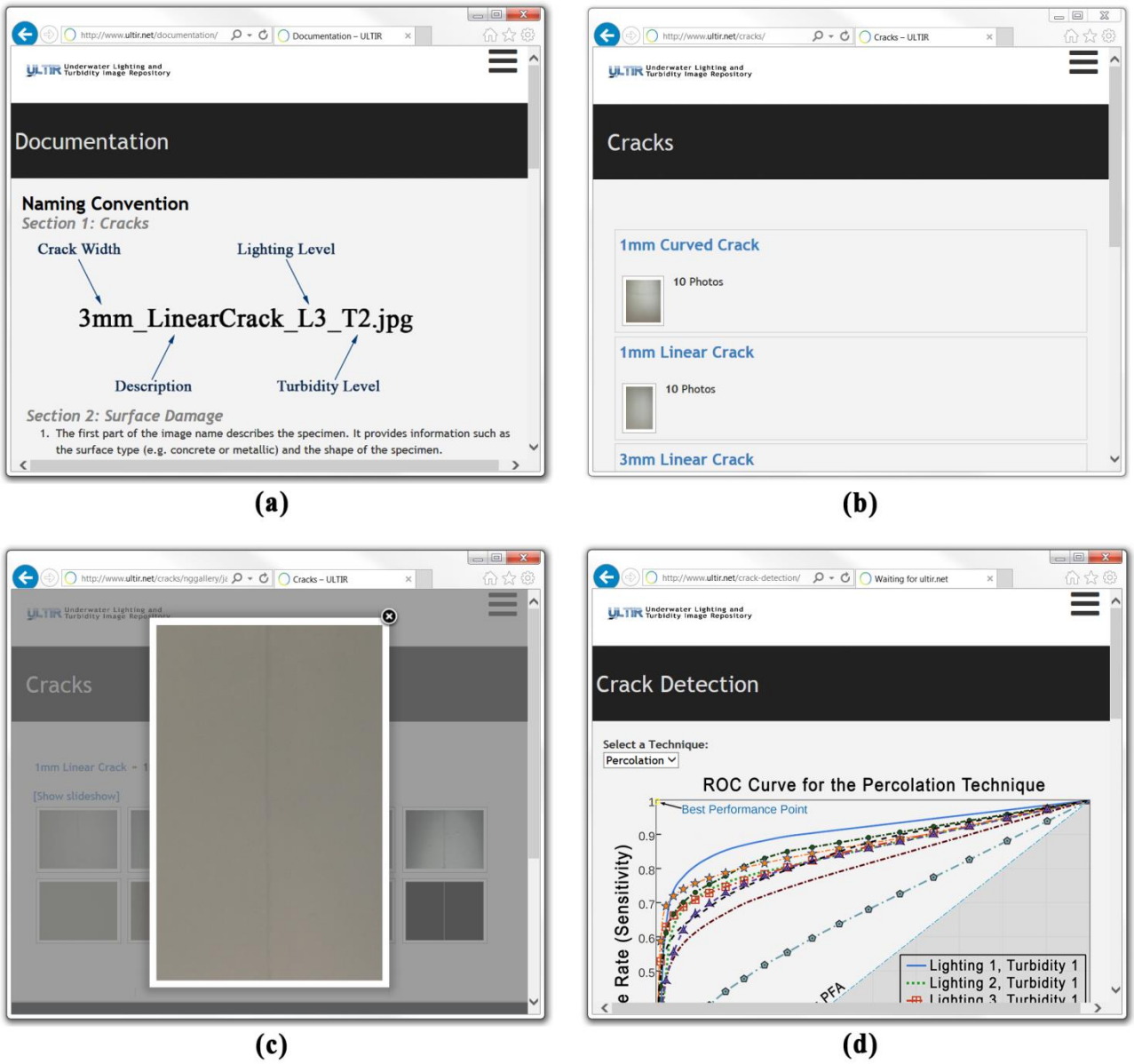

Figure 10. (a) ULTIR documentation and user guide, (b) Crack page which lists the specimens and nature of the crack, (c) sample image from the crack part of the repository captured in low light and high turbidity, and (d) Technique evaluation page.

The documentation and user guide, shown in Figure 10(a), explains the naming convention and advices how to navigate through the repository. The naming convention was adopted to succinctly convey important information such as the specimen description, turbidity level and lighting level. Figure 10(b) presents an overview of all images in each category of the repository, in this case, the crack repository, and reveals details about the nature of the crack such as the crack width. Clicking on any thumbnail in this list will bring up the full assortment of associated images for each turbidity and lighting level, along with a binary

control image which shows the damaged region in white and the undamaged in black, as shown in Figure 10(c). Metadata such as the aperture, focal length and ISO remain embedded in the images. The technique evaluation page is shown in Figure 10(d). The page allows inspectors to identify and adopt techniques that work well for a given situation, as well as allowing algorithm developers to compare their techniques against existing state-of-the-art.

3.0 Application of Image based Techniques

This section compares some of the main algorithmic approaches related to each category of ULTIR, namely; crack detection, surface damage detection, and 3D shape recovery. The purpose of this section is to characterize the performance of image algorithms under a given set of environmental conditions and to identify state-of-the-art methods. A number of algorithms are applied to the representative datasets shown in Figure 5-7. The performance of submitted techniques are evaluated and ranked using Receiver Operating Characteristic (ROC) curves. The ROC curves offer a convenient way of characterising the performance of NDT methods under various environmental conditions (Rouhan and Schoefs, 2003) and have been expanded to image detection (Pakrashi et al., 2008).

For any NDT technique, the Detection Rate (DR) along with the accompanying Misclassification Rate (MCR), which are similar to the Probability of Detection (PoD) and Probability of False Alarm (PFA) in the field of probability space and decision theory, are determined by comparing the damaged regions detected with a visually segmented image. The visually segmented image is created by a human operator who must manually identify damaged regions in an image. This visually segmented image acts as the control as it is assumed it shows the true extent of damage. The visually segmented image only needs to be created when it is wished to gauge the performance levels of the technique under scrutiny.

The *DR* and *MCR* are represented as a percentage between 0% and 100%. Each (*MCR*,*DR*) pair formed a coordinate in the ROC space. The *DR* and *MCR* are defined as:

$$DR \approx \frac{Card(E)}{n_c} \quad \text{with} \quad E = \left\{ g \in \Im ; \gamma_g = 1 \right\} \tag{1}$$

$$MCR \approx \frac{Card(F)}{n} \quad \text{with} \quad F = \left\{ g \in \Im ; \gamma_g = -1 \right\} \tag{2}$$

where Card(.) indicates the cardinal of a particular set, $\Im = \{1, \ldots, n\}$, n is the total number of pixels in the image, n_c denotes the number of damaged pixels and γ_g is an instance label vector, where $\gamma_g = 1$ corresponds to correctly identified non-damaged pixels, i.e. true positives, and $\gamma_g = -1$ corresponds to incorrectly detected pixels and undetected damaged pixels, i.e. false negatives and false positives. *F* gathers situations of incorrectly detected pixels and undetected damaged pixels while *E* gathers the correctly detected ones.

The α-δ method is used to compare the performance of related techniques (Baroth et al., 2011, Schoefs et al., 2012a). This method relies on calculating the angle, α, and the Euclidean distance, *δ*, between the best performance point, defined as an ideal NDT technique with 100% detection and 0% misclassification rates and represented in the ROC space with coordinates (0,1) and the considered point to give a measure of the performance of the considered point. As this paper does not deal with risk analysis where the shape to the ROC acts as a key factor, only the delta, *δ*, parameter is required as a measure of performance. A low value for *δ* is indicative of a strong performance. While there has been recent spurts of interest around image processing and structural monitoring (Jahanshahi et al., 2009; Jahanshahi et al., 2011; Zaurin and Catbas, 2011; Le Blanc et al., 2013; Kohut et al., 2013; Kim et al, 2013), comprehensive databases have not been developed for such purposes and comprehensive studies have been seldom carried out despite the use of such techniques for a significantly longer period of time (Patsias and Staszewskiy, 2002).

3.1 Crack Detection

The pervasive nature of cracks and the tedious task of manually counting and measuring them has led to a growing interest in utilising image processing based techniques to automate the detection process. Once cracks have been automatically detected from the inspection imagery, physical properties of the identified cracks, such as the length and the width, can be extracted. These physical properties need to be converted from pixel dimensions to real world units (e.g. millimetres) in order to be meaningful to engineers and to be of any value for subsequent analysis tasks. This requires establishing a scaling factor that relates pixel dimensions to real world units. Possible ways of obtaining a scaling factor include practical approaches such as placing an object of known dimensions alongside the crack, or by using a calibrated stereo system to obtain full 3D information (O'Byrne et al. 2016). A number of techniques have been devised which are capable of identifying crack-like features which are characterized by their narrow shape and lower brightness in comparison to the surroundings. These include a percolation based method (O'Byrne et al., 2014a), eigenvalue analysis of the Hessian (Frangi et al., 1998), Kirsch templates (Kirsch, 1971), neural networks (Choudhary and Dey, 2012) and statistical filters (Sinha and Fieguth, 2006). This paper applies the first three of these methods to the 1 mm controlled crack data set (as shown in Figure 5.5) from the repository to investigate the effects of changing turbidity and lighting levels on the detection accuracy for each technique.

In practice, the input for these crack detection methods is an image, or a batch of images, that feature visible cracks. The output is a binary image (black and white image) in which the white pixels denote pixels that likely correspond to cracks, while black pixels correspond to the background. Crack detection methods employ various approaches to identify and isolate cracks, some of which are described below. The percolation method is based on tracing out dark pixels within a fixed-sized window, or sub-image, starting at the centre point of the window, and spreading out until the boundary of the window is reached. The resulting pattern

of dark pixels is analysed. Cases where narrow or linear patterns are traced out are indicative of cracks, while irregular or radial diffusion patterns typically correspond to the non-cracked background. The eigenvalue analysis of the Hessian method detects narrow crack-like paths by calculating the direction of smallest curvature where there is minimum change in intensity, which is usually along the crack path. The Kirsch templates method detects line-like objects using spatial filtering involving templates orientated in eight different directions followed by thresholding.

The results for each of these techniques are shown in Figure 11, the performance levels are summarized in Table 2.

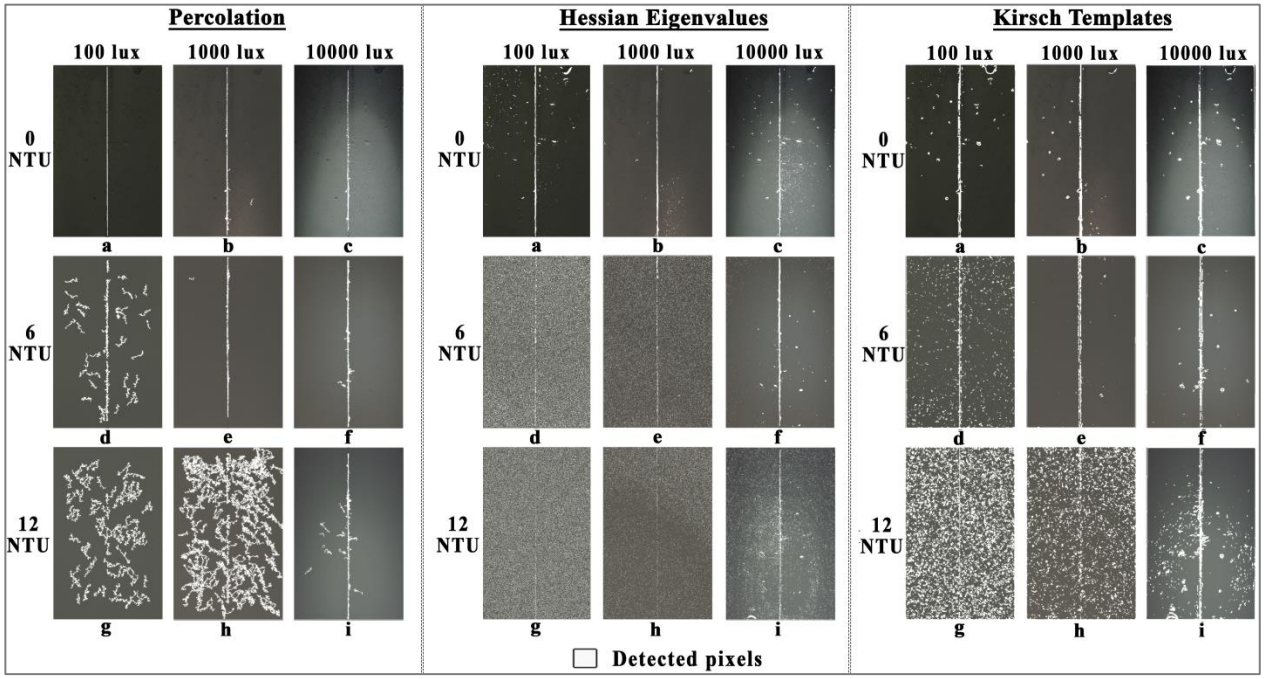

Figure 11. Detected cracks corresponding to the images in Figure 5.

Table 2. Performance of the crack detection techniques.

Image	Condition	Percolation			Hessian Eigenvalues			Kirsch Templates		
		(DR)	(MCR)	δ	(DR)	(MCR)	δ	(DR)	(MCR)	δ
a	Lighting 1, Turbidity 1	94.9%	0.7%	0.05	96.7%	2.3%	0.04	92.3%	5.3%	0.09
b	Lighting 2, Turbidity 1	91.5%	0.8%	0.09	91.6%	1.2%	0.09	94.1%	4.0%	0.07
c	Lighting 3, Turbidity 1	81.7%	0.7%	0.18	90.7%	3.9%	0.10	96.8%	5.3%	0.06
d	Lighting 1, Turbidity 2	75.0%	3.9%	0.25	58.6%	50.2%	0.65	67.9%	13.2%	0.35
e	Lighting 2, Turbidity 2	88.0%	1.0%	0.12	54.6%	34.1%	0.57	79.4%	4.5%	0.21
f	Lighting 3, Turbidity 2	94.1%	1.2%	0.06	95.4%	7.7%	0.09	92.7%	4.8%	0.09
g	Lighting 1, Turbidity 3	21.5%	8.4%	0.79	44.6%	48.7%	0.74	56.4%	54.2%	0.70
h	Lighting 2, Turbidity 3	54.8%	21.7%	0.50	25.2%	29.8%	0.81	55.3%	36.1%	0.57
i	Lighting 3, Turbidity 3	91.7%	1.7%	0.08	83.7%	24.6%	0.30	92.5%	11.7%	0.14

The ROC curves are generated by applying the techniques to all of the cracked specimens in the repository multiple times, each time adjusting the considered technique's critical parameters, and comparing the resulting detected cracks with the corresponding ground truth data in order to obtain a set of sensitivity and specificity pairs.

The ROC curves for all of the specimens under a certain lighting and turbidity condition are then averaged to produce a curve representative of the overall performance of a particular technique at that lighting and turbidity condition. This results in 9 curves for each technique, as shown in Figure 12:

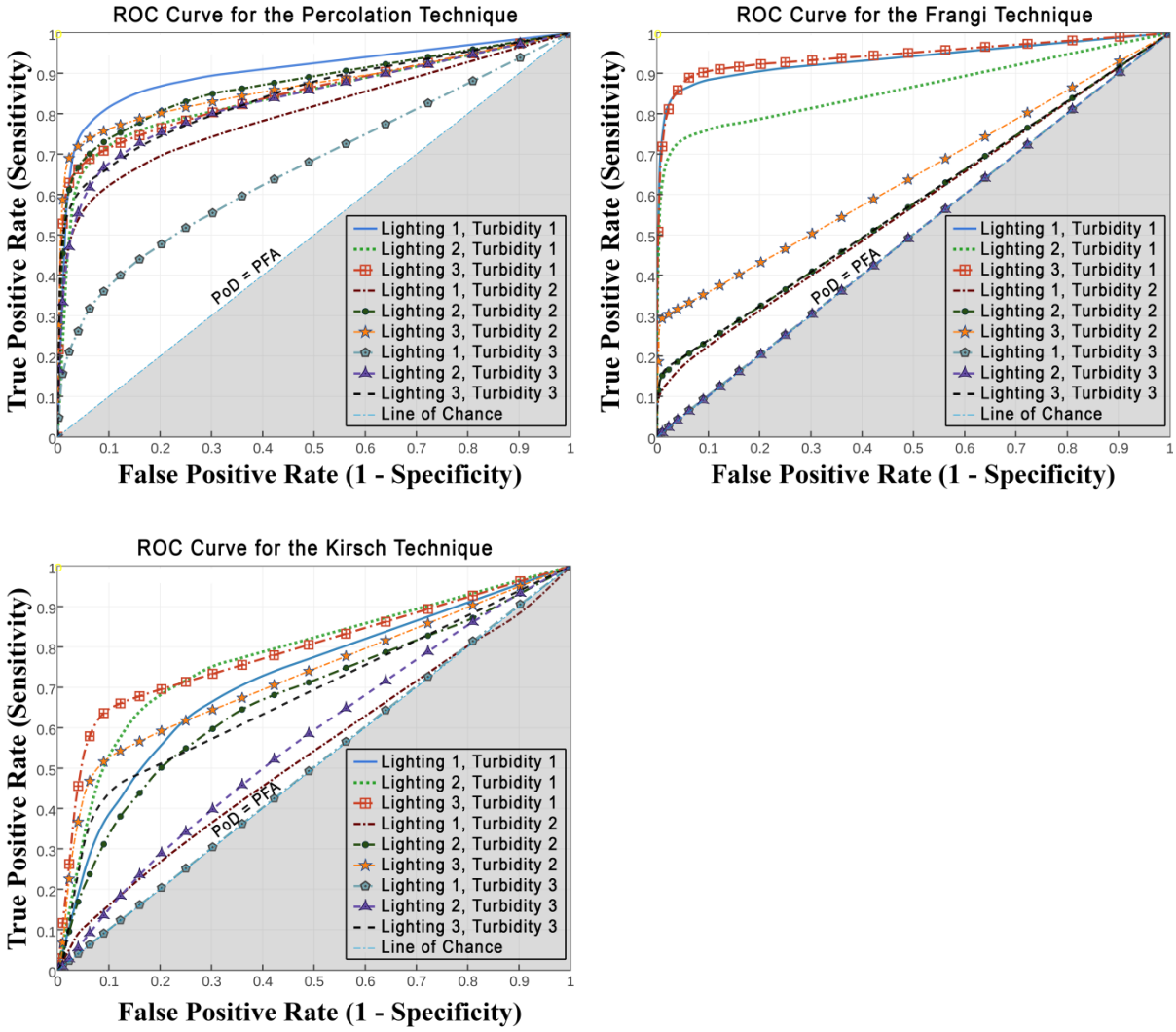

Figure 12. Evaluation of crack detection techniques through the use of ROC Curves.

It may be observed from the detected cracks in Figure 11 that each technique performs quite well for images that feature clear and sufficiently lit scenes. Unsurprisingly, the performance deteriorates when the turbidity levels increase. In the worst visibility conditions - low light and high turbidity (Figure 5g) - all of the techniques produce poor results suggesting the adoption of image based crack detection approaches under these conditions is not practical. However, the results show that having high lighting can mitigate the effects of high turbidity. In these situations, the increased absorption and diffusion in turbid water limits the formation of a bright spot which would otherwise impair detection.

Adopting the α-δ method allowed for a clear comparison between various (DR,MCR) pairs. Analysis of the δ parameter in Table 2 reveals that the percolation based method performs the best for the medium turbidity level. The cracks are well delineated and there are relatively few misclassified pixels. The eigenvalue analysis of the Hessian method produces a high misclassification rate at the lower light levels, while the Kirsch templates method also produces a lot of small spurious regions.

Overall, there is relatively little difference in terms of detection accuracy for the techniques considered. Instead, the visibility conditions have by far the greatest influence on the output.

3.2 Surface Damage

Most image processing based damage detection algorithms consist of segmentation followed by subsequent classification of the segmented regions. Ideally, the segmentation methodology should identify and accurately define all regions of interest in an image whilst minimizing the inclusion of extraneous regions. In reality, perfect segmentation is difficult to achieve given the inherent chromatic and luminous complexities encountered in natural scenes. Image processing based techniques include colour intensity based methods and texture analysis based methods. Naturally, the techniques in each group are suited to different applications. The effectiveness of colour based segmentation algorithms and texture based segmentation algorithms will vary according to the surface and damage type under consideration as certain damages are more separable from the undamaged surface based on either their colour or texture attributes. This section assesses the performance of two colour based methods, REMPS (O'Byrne et al., 2014b) and Otsu's thresholding (Otsu, 1979); along with a texture analysis based technique (O'Byrne et al., 2013b) previously proposed in the domain of NDT.

Colour based segmentation techniques can be either contextual or non-contextual. Non-contextual techniques such as Otsu's method do not take into account any spatial relationships

between pixels in an image. Each pixel is considered individually and segmentation is performed on the basis of some global attribute such as the pixel intensity value. Non-contextual techniques typically have lower algorithmic complexity, wider availability, and faster computational speeds compared to contextual techniques, yet they can still achieve good results as evident from the performance of Otsu's method presented in Table 3. REMPS, which operates by classifying whole regions, is an example of a contextual technique that takes into account spatial relationships amongst neighbouring pixels. If a contextual relationship is an important factor for segmenting a particular image, than non-contextual techniques will have limited success compared to contextual techniques.

These techniques are applied to the imagery in Figure 6 and the results are shown in Figure 13, the performance levels are quantified in Table 3.

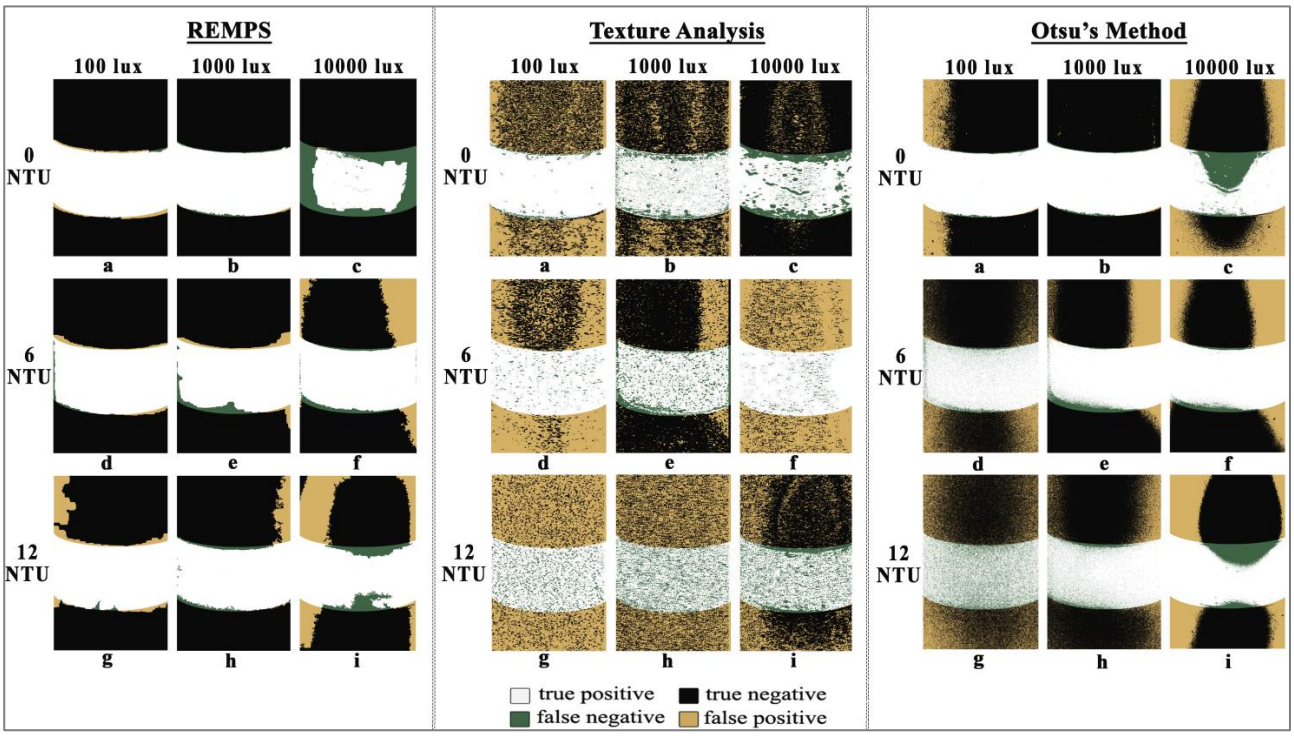

Figure 13. Detected damage corresponding to the images in Figure 6.

Table 3. Performance of the damage detection techniques.

Image	Condition	REMPS			Texture Analysis			Otsu's Method		
		(DR)	(MCR)	δ	(DR)	(MCR)	δ	(DR)	(MCR)	δ

a	Low Light, Low Turbidity	99.5%	2.1%	0.02	97.2%	49.3%	0.49	98.1%	14.8%	0.15
b	Medium Light, Low Turbidity	97.5%	1.2%	0.03	89.9%	33.1%	0.35	97.5%	1.2%	0.03
c	High Light, Low Turbidity	60.3%	14.2%	0.42	75.5%	15.8%	0.29	76.5%	39.8%	0.46
d	Low Light, Medium Turbidity	98.1%	2.5%	0.03	98.0%	51.3%	0.51	95.7%	22.2%	0.23
e	Medium Light, Medium Turbidity	91.9%	5.7%	0.10	82.6%	20.6%	0.27	92.6%	18.9%	0.20
f	High Light, Medium Turbidity	93.3%	16.7%	0.18	98.8%	61.5%	0.61	94.1%	24.6%	0.25
g	Low Light, High Turbidity	99.2%	7.0%	0.07	90.1%	61.8%	0.63	83.5%	37.5%	0.41
h	Medium Light, High Turbidity	93.4%	6.9%	0.10	83.8%	58.0%	0.60	93.9%	20.9%	0.22
i	High Light, High Turbidity	85.4%	24.0%	0.28	76.8%	34.0%	0.41	82.9%	30.7%	0.35

ROC curves for each technique under the varying lighting and turbidity conditions are shown in Figure 14.

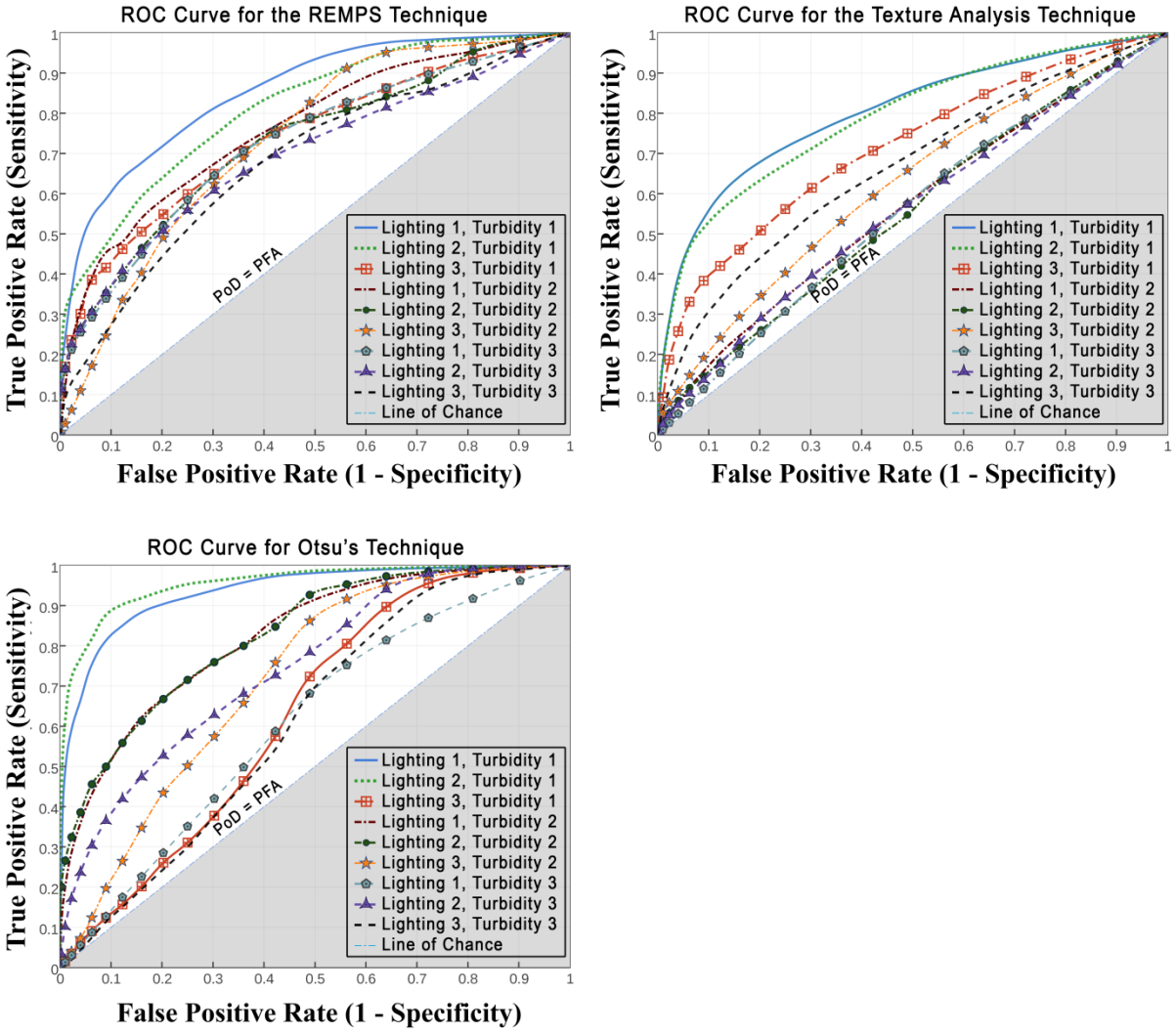

Figure 14. Evaluation of damage detection techniques through the use ROC curves.

These ROC curves are generated by applying the techniques to all of the specimens in the surface damage section of ULTIR multiple times, each time adjusting the considered

technique's critical parameters, and comparing the resulting detected damaged region with the corresponding ground truth data in order to obtain a set of sensitivity and specificity pairs. The ROC curves for all of the specimens under a certain lighting and turbidity condition are then averaged to produce a curve representative of the overall performance of a particular technique at that lighting and turbidity condition. This results in 9 curves for each technique.

It may be noted from these results that REMPS was quite robust - it displayed less sensitivity to the input conditions as evidenced by the close bunching of ROC curves across all lighting and turbidity levels in Figure 14. Moreover, it proved effective at locating the presence of damage as well as accurately defining the shape and size of damaged regions. The success of Otsu's method may be explained by the fact that the damaged region for the specimen is characterized by one single colour which is sufficiently distinct from the background. Both Otsu's method and texture analysis had many small spurious regions unlike REMPS which produced a 'cleaner' and more homogenous detection.

The texture based method was effective at locating the presence of damage as may be observed from Figure 13; however it did not perform as well as the colour based methods at defining the extent of damage which resulted in poor DR, MCR and δ values in Table 3. The exception to this was for the low turbidity and high light image (Figure 6c) where the texture analysis outperformed the colour based methods. While the high light and the shiny metallic surface created luminous complexities that misled the colour based methods, the high light illuminated and brought out some of textural properties of the surface which benefitted the texture analysis technique.

While the visibility conditions had a major impact on the detection of cracks, the performance of surface damage techniques do not rapidly decline with deteriorating visibility condition. Instead, the results indicate that choice of technique is more critical.

3.3 3D Shape Recovery using Stereo Vision

Stereo vision is the process of recovering depth from images, typically taken by two cameras that are separated by a horizontal distance. The recovery of 3D structure using stereo vision requires two sub-problems to be solved: the correspondence problem, in which image points corresponding to the same object are matched, and the reconstruction problem, in which the matched image points are reconstructed into 3D information (Ho & Chung, 1997).

The correspondence problem is often difficult because of ambiguous correspondences that can lead to false matches. This is especially problematic for uniform surface types whereby the lack of distinct features causes a high number of vague matches. Such scenarios are particularly likely to arise in underwater settings due to the poor visibility conditions. Additionally, another problem arises when correspondence cannot be reliably established because a region is occluded in one of the images. The correspondence problem has been extensively researched and a wide range of stereo matching algorithms have been proposed. (Scharstein and Szeliski, 2002) provide a taxonomy of the most notable algorithms.

This section compares the performance of three distinct types of stereo correspondence algorithms. The first is a Belief Propagation (BP) Markov Random Field (MRF) method (Sun et al., 2002), which is a hierarchical (coarse-to-fine) algorithm that operates on an image pyramid, where results from coarser levels are used to constrain a more local search at finer levels. The Markov Random Field model takes into account the differences between pixel intensity values between corresponding points and the spatial relationship between the horizontal disparities. The disparity relates how far each point is from the camera, i.e. its depth in the scene. The goal is to find a piecewise smooth horizontal disparity map consistent with the observed data which minimizes total energy. This method promotes a smooth disparity map as it penalizes cases where neighbouring pixels have different disparity values.

The second method is based on the well-known Birchfield and Tomasi (1998) matching cost that is robust to image sampling. This method compares each pixel in the reference

image against a linearly interpolated function of the other image. It does not rely on any smoothness constraints, but rather, the disparity is computed by selecting the minimal (winning) aggregated value at each pixel.

The final method is a region based stereo matching algorithm (Alagoz, 2008), which operates by semi-global error energy minimization by smoothing functions. This method chooses a root point in a given region and then grows that point while the energy function remains equal or less than a certain value. Otherwise, a new root point is selected and separate region starts growing from there. This algorithm employs the Sum of Absolute Differences (SAD) as a similarity measure.

These techniques are applied to the stereo imagery featuring a concrete cube in Figure 7 and the resulting disparity maps, or depth maps, are shown in Figure 15. A low Normalized Root-Mean-Square Error (NRMSE) score is indicative of a strong performing stereo correspondence technique as it corresponds to cases where there is a small deviation between the computed disparity map and the ground truth disparity map.

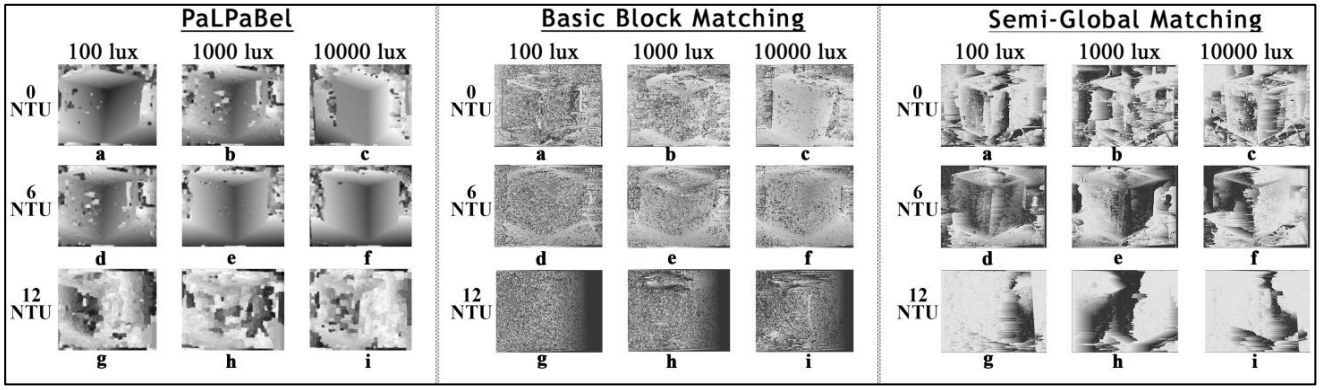

Figure 15. Disparity maps for the stereo imagery in Figure 7.

The aggregated NRMSE scores for all of the specimens in the repository under each lighting and turbidity are illustrated in the bar chart in Figure 16. A description of the specimens is provided in Table 4.

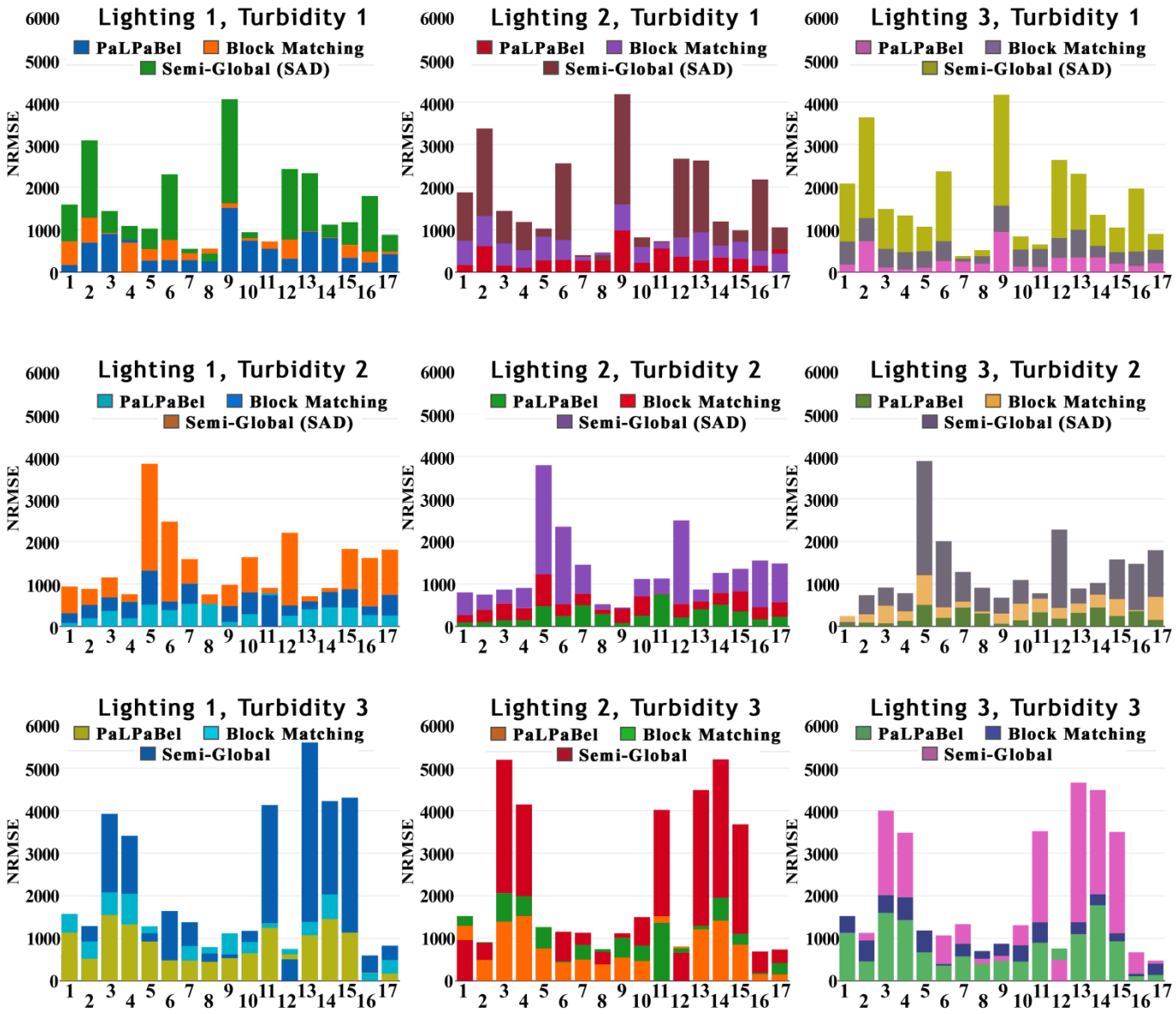

Figure 16 The aggregated NRMSE scores for all of the specimens in the repository under each lighting and turbidity.

Table 4. Description of specimens

1 Rough Concrete Cube Planar	**2** Rough Concrete Cube Diagonal	**3** Smooth Concrete Cube Planar	**4** Smooth Concrete Cube Diagonal	**5** Concrete Cylinder	**6** Concrete Sphere
7 Metal Case Diagonal	**8** Metal Case Planar	**9** Metal Cube Planar	**10** Metal Cube Diagonal	**11** Metal Cylinder	**12** Metal Sphere
13 Plastic Cube Planar	**14** Plastic Cube Diagonal	**15** Plastic Cylinder	**16** Plastic Sphere	**17** Rubber lattice	

Results show a marked variation across the range of specimens in ULTIR. Specimens, such as the rough concrete cube (specimen no. 1) and the aged metallic case (specimen no. 8), facilitate more reliable stereo matching due to the richly textured nature of these surfaces

This is in contrast with specimens that have more uniform and homogeneous surfaces, such as the plastic cylinder (specimen no 15), whereby the lack of unique features is seen to inhibit the performance of the stereo algorithms.

It may be observed from the disparity maps in Figure 15 that the Belief Propagation technique produces the best results in most cases. Given that a key feature of this technique is that it explicitly accounts for smoothness between neighbouring pixels, a reasonable interpretation would be that algorithms also incorporating some form of smoothness term would fare better when applied to images in the ULTIR repository, and ultimately, when performed on images obtained from real-world underwater inspection. The usefulness of a smoothness term appears to be even more pronounced when operating at higher turbidity and light levels at which point some of the surface details become washed out. However, in the worst conditions (i.e. low light and high turbidity), the performances from all techniques are poor. This indicates that the operating limit has been reached.

Further analysis of the results indicates that the choice of similarity measure is an important factor in terms of the performance of a technique. The SAD similarity measure, which is used for the semi-global technique, is not as effective as the Zero-Mean Normalized Cross Correlation (ZNCC) score, which is widely acknowledged as being a robust similarity measure in the literature (Krattenthaler et al., 1994; O'Byrne et al., 2015). These results demonstrate that a good understanding of the operating conditions, along with a careful and considered choice of image processing algorithm, is important for maximising the information that can be obtained from image analysis. This analysis could only be achieved with a resource such as ULTIR.

Conclusions

Image processing based methods are increasingly being used to provide a source of quantitative information when inspecting offshore and marine structures. In order for inspectors to fully exploit the potential of image processing based methods, it is important that they are aware of the relationship between underwater visibility, which is largely governed by the on-site lighting and turbidity levels, and the ability of image processing algorithms to successfully identify and quantify instances of damage. This paper describes an open-source Underwater Lighting and Turbidity Image Repository (ULTIR). The aim of ULTIR is lessen the uncertainty introduced by the onsite visibility conditions, thereby enabling inspectors to make informed decisions when using image processing methods. The resource also assists researchers when developing and evaluating new or existing image processing algorithms intended for underwater application.

Image processing algorithms in the domain of SHM have been developed for numerous purposes, such as crack detection, surface damage detection, and 3D shape recovery using stereo-vision. A diverse array of algorithms have been devised for each purpose, however, there exists no standardized approach for choosing a technique that is best-equipped to deal with the challenging visibility conditions. This paper carries out an initial ranking of several image based methods using the imagery and ground-truth data from the repository. The results show that turbidity has a major effect on the accuracy of the non-destructive techniques. While increasing turbidity adversely impacts technique accuracy, the relationship between the illumination and the technique accuracy is more complicated. A dark, poorly lit scene is clearly unfavourable for the task of detecting damage; however, too much lighting also creates problems, such as bright spots on the surface of objects, which obscures detail. While these general trends emerged from this study, the illumination and turbidity levels have a combined effect on the technique accuracy and must therefore be considered jointly.

The accuracy of the non-destructive techniques also depends on the particular methodology followed by each technique to assess the damage. For instance, contextual damage detection techniques (i.e. those that take into account spatial relationships between neighbouring pixels in an image), were found to perform better than non-contextual techniques (i.e. those that classify a pixel as being damaged or non-damaged by only considering that pixel independent of nearby pixels). This was especially true in reduced visibility conditions where non-contextual techniques were highly sensitive to noise arising from the presence of turbidity and the complex underwater light-field. With this in mind, the choice of method is an important factor. It was found that the performances of some techniques diminish markedly in harsh lighting and turbidity conditions in comparison with other techniques. For example, the Frangi technique for detecting cracks is the best performer in low turbidity conditions; however, it performs very poorly in high turbidity conditions. Meanwhile, the percolation based crack detection method achieves a much more consistent performance across the full range of turbidity levels. The same can be said for Otsu's method and the REMPS technique which are used for detecting surface damage. Otsu's method performs the best in low turbidity conditions, while it is amongst the worst performers in high turbidity conditions. REMPS, on the other hand, remains a consistently good performer, with only minor performance deterioration with increasing turbidity. Additionally, the results reveal that in many cases high turbidity can be partially mitigated by having bright light source.

Image based methods undoubtedly have value as a quantitative inspection tool, however, the performance of these methods are often hampered by the poor visibility and complex underwater light field. This large and well-annotated dataset paves the way for researchers to develop data-driven algorithms, such as deep learning techniques, that may be more robust to the underwater conditions. Such algorithms have already attracted significant interest in other fields owing to the high performances that they can achieve; however, they have not had an

impact in the domain of underwater damage detection as of yet, largely due to the lack of available training data. It is expected that ULTIR will continue to evolve as new imagery is added and more techniques are benchmarked.

ACKNOWLEDGEMENTS

The authors wish to thank the Irish Research Council for Science, Engineering and Technology (IRCSET) and CAPACITES/IXEAD society for providing financial and practical support. The authors acknowledge Marine Research Energy Ireland (MaREI), grant no. 12/RC/2302, a Science Foundation Ireland (SFI) supported project.

References

Agin, G. J. (1980). Computer Vision Systems for Industrial Inspection and Assembly. *Computer*, 13, 11-20. doi:10.1109/MC.1980.1653613

Alagöz, B. B. (2008). Obtaining depth maps from color images by region based stereo matching algorithms. *OncuBilim Algorithm And Systems Labs*, 8, 1-13. Retrieved from https://arxiv.org/ftp/arxiv/papers/0812/0812.1340.pdf

Anderson, M. R. (1987). Nondestructive testing of offshore structures. *NDT international*, 20, 17-21. doi:10.1016/0308-9126(87)90368-3

Barnouin, B., Lemoine, L., Dover, W. D., & Rudlin, J. (1993). Underwater inspection reliability trials for offshore structures. *Proceedings Of The International Conference On Offshore Mechanics And Arctic Engineering*, 3, 883-883.

Baroth, J., Breysse, D., & Schoefs, F. (2011). *Construction Reliabilit*. Hoboken: NJ Wiley.

Bianco, G., Gallo, A., Bruno, F., & Muzzupappa, M. (2013). A comparative analysis between active and passive techniques for underwater 3D reconstruction of close-range objects. *Sensors*, 13, 11007-31. doi:10.3390/s130811007

Birchfield, S., & Tomasi, C. (1998). Pixel dissimilarity measure that is insensitive to image sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 401-406. doi:10.1109/34.677269

Boukinda Mbadinga, M. L., Schoefs, F., Quiniou-Ramus, V., Birades, M., & Garretta, R. (2007). Marine growth colonization process in Guinea Gulf: Data analysis. *Journal of Offshore Mechanics and Arctic Engineering*, 129, 97-106. doi:10.1115/1.2355518

Bruno, F., Bianco, G., Muzzupappa, M., Barone, S., & Razionale, A. (2011). Experimentation of structured light and stereo vision for underwater 3D reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66, 508-518. doi:10.1016/j.isprsjprs.2011.02.009

Busby, F. R. (1979). Underwater Inspection/Testing/Monitoring of Offshore Structures. *Ocean Engineering*, 6, 355-491. doi:10.1016/0029-8018(79)90006-4

Dana, K. J., Nayar, S. K., Van Ginneken, B., & Koenderink, J. J. (1997). Reflectance and texture of real-world surfaces. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 17-19 June 1997, Los Alamitos, CA, USA. 151-157.

Dirksen, J., Clemens, F. H. L. R., Korving, H., Cherqui, F., Le Gauffre, P., Ertl, T., Plihal, H., Müller, K., & Snaterse, C. T. M. (2013). The consistency of visual sewer inspection data. *Structure and Infrastructure Engineering*, 9, 214-228. doi:10.1080/15732479.2010.541265

Ebrahimkhanlou, A., Farhidzadeh, A., & Salamone, S. (2016). Multifractal analysis of crack patterns in reinforced concrete shear walls. *Structural Health Monitoring*, 2016, 15, 81-92. doi:10.1177/1475921715624502

Estes, A., & Frangopol, D. (2003). Updating bridge reliability based on bridge management systems visual inspection results. *Journal of Bridge Engineering*, 8, 374-382. doi:10.1061/(ASCE)1084-0702(2003)8:6(374)

Graham, A. A. (1990). Siltation of stone-surface periphyton in rivers by clay-sized particles from low concentrations in suspention. *Hydrobiologia*, 199, 107-115. doi:10.1007/BF00005603

GVSU (Grand Valley State University). 2013. Instructor's Manual – Turbidity. Available online at: http://www.gvsu.edu/wri/education/instructor-s-manual-turbidity-10.htm, [Accessed 10 December 2016].

Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, 3**,** 610-621. doi:10.1109/TSMC.1973.4309314

Hearn, G., & Testa, R. B. (1991). Modal Analysis for Damage Detection in Structures. *Journal of Structural Engineering*, 117, 3042-3063. doi:10.1061/(ASCE)0733-9445(1991)117:10(3042)

Ho, P. K., & Chung, R. (1997). Stereo motion that complements stereo and motion analyses. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, June 1997, Peurto Rico. 213-218. doi:10.1109/CVPR.1997.609322

Iyer, S., & Sinha, S. K. (2006). Segmentation of pipe images for crack detection in buried sewers. *Computer-Aided Civil and Infrastructure Engineering*, 21, 395-410. doi:10.1111/j.1467-8667.2006.00445.x

Jahanshahi, M. R., Masri, S, F., & Sukhatme, G. S. (2011). Multi-image stitching and scene reconstruction for evaluating defect evolution in structures. *Structural Health Monitoring*, 10, 643-657. doi:10.1177/1475921710395809

Jahanshahi, M. R., Kelly, J, S,. Masri, S. F., & Sukhatme, G. S. (2009). A survey and evaluation of promising approaches for automatic image-based defect detection of bridge structures. *Structure and Infrastructure Engineering*, 6, 455-486. doi:10.1080/15732470801945930

Kim, S. W., Jeon, B. G., Kim, N. S., & Park, J. C. (2013). Vision-based monitoring system for evaluating cable tensile forces on a cable-stayed bridge. *Structural Health Monitoring*, 12, 440-456. doi:10.1177/1475921713500513

Kohut, P., Holak, K., Uhl, T., Ortyl, Ł., Owerko, T., Kuras, P., & Kocierz, R. (2013). Monitoring of a civil structure's state based on noncontact measurements. *Structural Health Monitoring*, 12, 411-429. doi:10.1177/1475921713487397

Komorowski, J. P., & Forsyth, D. S. (2000). Role of enhanced visual inspections in the new strategy for corrosion management. *Aircraft Engineering and Aerospace Technology*, 72, 5-13. doi: 10.1108/00022660010308615

Krattenthaler, W., Mayer, K. J., & Zeiler, M. (1994). Point correlation: A reduced-cost template matching technique. *IEEE Int. Conf. on Image Processing (ICIP 1994)*, September 1994, Austin, Texas, USA, 208–212. doi:10.1109/ICIP.1994.413305

Lawlor, A., Torres, J., O'Flynn, B., Wallace, J., & Regan, F. (2012). DEPLOY: A long term deployment of a water quality sensor monitoring system. *Sensor Review*, 32, 29-38. doi:10.1108/02602281211197125

LeBlanc, B., Niezrecki, C., Avitabile, P., Chen, J., & Sherwood J. (2013). Damage detection and full surface characterization of a wind turbine blade using three-dimensional digital image correlation. *Structural Health Monitoring*, 12, 430-439. doi:10.1177/1475921713506766

Lee, B. Y., Kim, Y. Y., Yi, S. T., & Kim, J. K. (2013). Automated image processing technique for detecting and analysing concrete surface cracks. *Structure and Infrastructure Engineering*, 9, 567-577. doi:10.1080/15732479.2011.593891

Mei, X., Gunaratne, M., Lu, J. J., & Dietrich, B. (2004). Neural network for rapid depth evaluation of shallow cracks in asphalt pavements. *Computer-Aided Civil and Infrastructure Engineering*, 19, 223-230. doi:10.1111/j.1467-8667.2004.00350.x

Nishikawa, T., Yoshida, J., Sugiyama, T., & Fujino, Y. (2012). Concrete crack detection by multiple sequential image filtering. *Computer-Aided Civil and Infrastructure Engineering*, 27, 29-47. doi:10.1111/j.1467-8667.2011.00716.x

O'Byrne, M., Pakrashi, V., Ghosh, B., & Schoefs, F. (2016). Evaluation of Camera Calibration Techniques for Quantifying Deterioration. *Proceedings of the Civil Engineering Research in Ireland Conference (CERI 2016),* August 29-30 2016. Galway, Ireland

O'Byrne, M., Ghosh, B., Pakrashi, V., & Schoefs, F. (2013a). Effects of turbidity and lighting on the performance of an image processing based damage detection technique. *11th International Conference on Structural Safety and Reliability, ICOSSAR 2013*, 16 – 20 June 2013, New York. 2645-2650. doi:10.1201/b16387-382

O'Byrne, M., Ghosh, B., Pakrashi, V., & Schoefs, F. (2014a). Effects of Turbidity and Lighting on an Image Processing based Crack Detection Technique. *Civil Engineering Research Ireland (CERI) Conference*, 28-29 August 2014, Belfast, Northern Ireland. doi:10.13140/2.1.3124.8162

O'Byrne, M., Schoefs, F., Pakrashi, V., & Ghosh, B. (2014b). Regionally enhanced multi-phase segmentation technique for damaged surfaces. *Computer-Aided Civil and Infrastructure Engineering*, 29, 644-658. doi:10.1111/mice.12098

O'Byrne, M., Pakrashi, V., Schoefs, F., & Ghosh, B. (2014c). A Comparison of Image Based 3D Recovery Methods for Underwater Inspections. *EWSHM - 7th European Workshop on Structural Health Monitoring*, 8 – 11 July 2014, Nantes, France.

O'Byrne, M., Schoefs, F., Ghosh, B., & Pakrashi, V. (2013b). Texture analysis based damage detection of ageing infrastructural elements. *Computer-Aided Civil and Infrastructure Engineering*, 28, 162-177. doi:10.1111/j.1467-8667.2012.00790.x

O'Byrne, M., Ghosh, B., Schoefs, F., O'Donnell, D., Wright, W., & Pakrashi, V. (2015). Acquisition and Analysis of Dynamic Responses of a Historic Pedestrian Bridge using Video Image Processing. *Institute of Physics, Journal of Physics, Conference Series*, 682, 012054-1-8. doi: 10.1088/1742-6596/628/1/012054

Omondi, B., Aggelis, D. G., Sol, H., & Sitters, C. (2016). Improved crack monitoring in structural concrete by combined acoustic emission and digital image correlation techniques. *Structural Health Monitoring*, 15, 359-378. doi:10.1177/1475921716636806

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-9, 62-6. doi:10.1109/TSMC.1979.4310076

Pakrashi, V., Schoefs, F., Memet, J. B., & O'Connor, A. (2008). ROC dependent event isolation method for image processing based assessment of corroded harbour structures. *Structure and Infrastructure Engineering*, 6, 365-378. doi:10.1080/15732470701718072

Patsias, S., & Staszewskiy, W. J. (2002). Damage detection using optical measurements and wavelets. *Structural Health Monitoring*, 1, 5-22. doi:10.1177/147592170200100102

Ramos, E. J. (1992). *Underwater inspection - the state of art: non-destructive* testing, Oxford, Elsevier.

Rouhan, A., & Schoefs, F. (2003). Probabilistic modeling of inspection results for offshore structures. *Structural Safety*, 25, 379-399. doi:10.1016/S0167-4730(03)00016-X

Rudlin, J. R. (1996). Reliability of inspection for fatigue cracks in offshore structures. *Inspection Reliability: State-of-the-Art (Digest No. 1996/178), IEE Colloquium on*, London, 1996, 6/1-6/3. doi: 10.1049/ic:19961032

Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47, 7-42. doi:10.1023/A:1014573219977

Schlyter, P. (2009). Radiometry and photometry in astronomy. Retrieved on August 10, 2015 at http://www.stjarnhimlen.se/comp/radfaq.html

Schoefs, F., Boéro, J., Clément, A., & Capra, B. (2012a). The αδ method for modelling expert Judgment and combination of NDT tools in RBI context: application to Marine Structures, Structure and Infrastructure Engineering: Maintenance, Management, Life-Cycle Design and performance (NSIE). *SMonitoring, Modeling and Assessment of Structural Deterioration in Marine Environments*, 8, 531-543. doi:10.1080/15732479.2010.505374

Schoefs, F., Abraham, O., & Popovics, J. (2012b). Quantitative evaluation of contactless impact echo for nondestructive assessment of void detection within tendon ducts. *Journal of Construction and Building Materials*, 37, 885-892. doi:10.1016/j.conbuildmat.2012.02.002

Sun, J., Shum, H. Y., & Zheng, N. N. (2002). Stereo Matching Using Belief Propagation. *Computer Vision - ECCV 2002: Proceedings of 7th European Conference on Computer Vision*, 28-31 May 2002, Copenhagen, Denmark, 510-524. doi:10.1007/3-540-47967-8_34

Valença, J., Dias-da-Costa, D., Gonçalves, L., Júlio, E., & Araújo, H. (2014). Automatic concrete health monitoring: assessment and monitoring of concrete surfaces. *Structure and Infrastructure Engineering*, 10, 1547-1554. doi:10.1080/15732479.2013.835326

Yazid, H., Arof, H., Yazid, H., Ahmad, S., Mohamed, A., & Ahmad, F. (2011). Discontinuities detection in welded joints based on inverse surface thresholding. *NDT & E International*, 44, 563-570. doi:10.1016/j.ndteint.2011.06.002

Zaurin, R., & Catbas, N. (2011). Structural health monitoring using video stream, influence lines, and statistical analysis. *Structural Health Monitoring*, 10, 309-332. doi:10.1177/1475921710373290