Defence Research and Development Canada
Recherche et développement pour la défense Canada

DRDC | RDDC
technologyscienceetechnologie

Development and Testing of an Underwater Optical Imaging System to Enable Unmanned Underwater Vehicle/Submarine Rendezvous

J. Giesbrecht
DRDC – Suffield Research Centre

G. D. Watt
DRDC – Atlantic Research Centre

Defence Research and Development Canada
Scientific Report
DRDC-RDDC-2018-R155
September 2018

Canada

IMPORTANT INFORMATIVE STATEMENTS

Abstract

This document describes the design and test of an optical position sensing system to enable autonomous rendezvous between a submarine and an Unmanned Underwater Vehicle (UUV). The vision system uses LED lighting on the UUV and a monocular camera system to predict the UUV position and pose relative to the camera. Important aspects include the use of a variety of computer vision algorithms changed adaptively over the course of the UUV approach, as well as the use of active control of the camera exposure to enable computer vision. Tests were conducted in harbour water from a barge facility at DRDC Atlantic in Bedford Basin, Nova Scotia using a mechanical ground truth system to verify performance. The system was found to operate in real-time at ranges to at least 13 m over a wide variety of relative positions, poses, and lighting conditions. Typical accuracy was on the order of less than 3% of range, and better than 5 degrees of accuracy in determination of UUV roll, pitch and yaw.

Significance for Defence and Security

It would be highly useful for military submarines to be able to deploy unmanned underwater vehicles to conduct various reconnaissance or minehunting tasks. Deploying them from the torpedo tubes is straightforward. However, recovering them without surfacing to recover data, recharge batteries, and dock for future use is difficult. The computer-vision system described in this report would allow an automated mechanical docking mechanism to capture a UUV while moving under the waves, improving the operational capabilities and mitigating risk for the submarine.

In particular, this report describes the algorithms, errors, and limitations of an optical position sensing system, all of which can be used to explore the viability of UUV docking through simulation.

Résumé

Le présent document vise à décrire la conception et l'essai d'un système optique de détection de la position pour permettre l'amarrage automatique d'un sous-marin à un véhicule sous-marin sans équipage (VSSE). Grâce à un éclairage à DEL à bord du VSSE et un système de caméra monoculaire, le système de vision permet de prévoir la position et l'orientation du VSSE par rapport à la caméra. Parmi les aspects importants du système, mentionnons l'utilisation de divers algorithmes de vision artificielle modifiés de manière adaptative à l'approche du VSSE, ainsi que le recours à un contrôle actif de l'exposition de la caméra visant à permettre une vision par ordinateur. RDDC Atlantique a procédé à des essais de rendement sur une barge, au moyen d'un système mécanique de réalité de terrain, dans les eaux portuaires du bassin de Bedford en Nouvelle Écosse. On a constaté que le système fonctionnait en temps réel jusqu'à une distance d'au moins 13 m, et ce, dans un large éventail de conditions d'éclairage, d'orientations et de positions relatives. La précision type s'est chiffrée à moins de 3 % de la distance et à plus de 5 degrés pour la mesure du roulis, du tangage et du lacet du VSSE.

Importance pour la défense et la sécurité

Il s'avérerait très utile que des VSSE puissent être déployés depuis des sous-marins militaires pour effectuer diverses opérations de reconnaissance ou de chasse aux mines. Leur déploiement est facile à réaliser à partir de tubes lance-torpilles. Toutefois, leur récupération pour en extraire les données, recharger les batteries et les amarrer pour un déploiement ultérieur se révèle difficile sans qu'on ait à faire surface. Le système de vision artificielle décrit dans le présent rapport permettrait à un mécanisme d'amarrage automatique de saisir le VSSE se déplaçant sous les vagues, ce qui améliorerait les capacités opérationnelles et réduirait les risques pour les sous marins. Le présent rapport vise tout particulièrement à décrire les algorithmes, les erreurs et les limites d'un système optique de détection de la position, éléments qui pourraient tous servir à étudier la viabilité de l'amarrage à un VSSE dans le cadre d'une simulation.

Table of contents

List of figures

List of tables

Acknowledgements

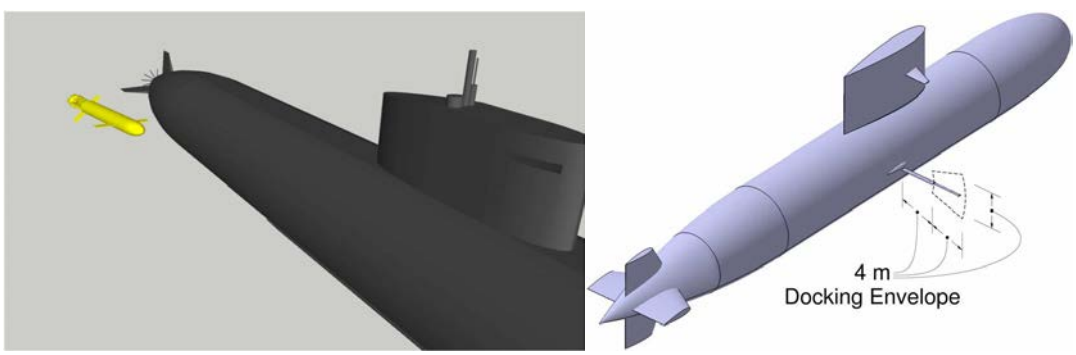
The authors very much appreciate the professional assistance of Roger Arsenault, Tomasz Lemczyk, Dang Phan, and Val Shepeta in developing the Pose Validation Rig and associated equipment and in running the experiments discussed herein.

1 Introduction

Modern diesel submarines operate in both the open ocean and littoral waters. Their operational capabilities can be expanded and risk mitigated with a deployable UUV (Unmanned Underwater Vehicle) capability that is fast, reliable, and has minimal operational limitations.

While launching UUVs is straightforward, recovering them to a submerged submarine is problematic, particularly in littoral waters in high sea states. These investigations were performed as part of a larger research project that aims to find an affordable and viable approach for docking UUVs onto submarines, involving sensing, capture and control capabilities necessary overcome relative motion between the vehicles caused by environmental disturbance.

Ideally, a recovery system would be autonomous, reliable, fast, and would allow recovery in littoral waters in high sea states. The overall system being designed in this project, described in [1], would have two stages: 1) a homing stage where the UUV moves towards the submarine and then holds course beside it, and 2) an active docking stage where the submarine is able to grab the UUV (Figure 1). The concept would be similar to the way aircraft are refueled in-flight.



(a) *UUV holding station beside a submarine for docking.*

(b) *One potential dock configuration.*

Figure 1: *UUV/Submarine active docking concept.*

Three types of remote sensing are planned to detect and track the UUV: acoustic, electromagnetic, and optical. To find the submarine during the homing stage, the UUV will use acoustics and electromagnetic sensing which are effective for medium and long range. However, during the final docking phase, while the UUV holds course, the docking mechanism will rapidly adapt to relative motions of the UUV for final capture. For this phase, a fast and accurate tracking method is required for shorter range sensing, on the order of 10–20 metres. This report documents the design, development, and test of a monocular visual optical tracking for this application. This concept, illustrated in Figure 2, would have a camera system on the dock end effector that would locate the relative position of light

sources on the UUV to allow for a feedback control loop to guide the dock onto the UUV's path for capture. The system is intended to estimate both position and pose over short ranges (up to 20 metres). The work is focused on determining the performance that can potentially be achieved with short range optical tracking in a realistic environment.

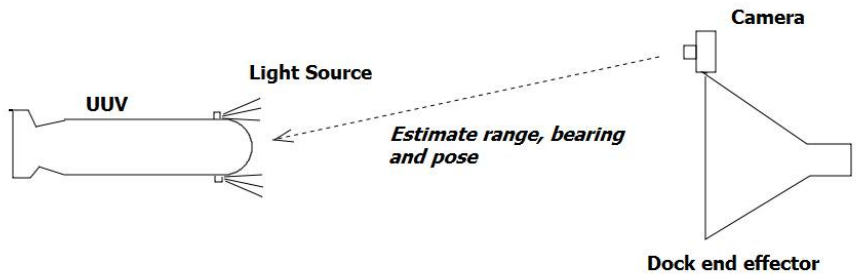


Figure 2: *Optical tracking to enable an active UUV dock.*

In a docking scenario, the relative position and pose of the UUV from the camera system will be used to guide the control algorithm for the active dock on the submarine. It is possible that a single light source could be used to "home-in" on the UUV for capture, but it is predicted that having more information on the location and heading of the UUV will allow for a more robust capture system.

The vision system has four major components: a digital camera to capture images, a computer-vision algorithm that pinpoints light sources from camera images, an exposure control algorithm that keeps the LED sources at an acceptable size in the images, and a pose estimation algorithm that uses the apparent positions of multiple LED lights. Along with the system design, results of experiments conducted in harbour water in October of 2016 are presented below.

The main contribution of this report is the computer-vision strategy and use of exposure control to enable docking using lights over the full range of imaging conditions, from far away to immediately in front of the camera. We also demonstrate the feasibility and accuracy of such an optical tracking system through experiments. This information can be used to develop a computer model of an optical position sensing system that can be used in a computer simulation to economically explore the viability of the above UUV docking concept.

2 Background

2.1 Challenges in Optical Imaging Underwater

Optical tracking provides fast data rates, and uses simple, inexpensive hardware with low power consumption. This is challenged by highly variable lighting and turbidity conditions resulting from the need to dock at variable depths (from 15 to 200 m), in daytime or nighttime, and in varying levels of clarity (the open ocean, littoral, harbour, or estuarial waters). Heavy concentrations of plankton and other particulate matter can make for murky or hazy waters where it is difficult to view objects at a distance.

For optical tracking, the absorption and scattering by the water itself, as well as by other particles such as silt or micro-organisms causes a reduction in the brightness and contrast of an object. This reduces detectable range and affects the availability of ambient light. The presence of marine life such as fish or jellyfish may also change the appearance of the target.

Jaffe et al. [2] discuss practical issues associated with underwater imaging. Backscatter results when light is reflected back toward a light source: it "limits the contrast of [a scene being illuminated by the observer] creating a veiling glow." Forward scatter occurs with light originating at or reflected from a target and results in blurring.

2.2 Approaches to Underwater Optical Tracking

Preliminary experiments were conducted by DRDC (reported in [3]), that discuss various options for underwater optical tracking, including:

- Laser Ranging—Laser radars that use time-of-flight measurements could possibly be used to find the relative position of a passive UUV [4, 5, 6]. This method is not ideal because it requires expensive and complicated hardware on the dock, and suffers from the return path problem where the emitted laser light needs to travel both to the target and back, and is therefore scattered and absorbed over twice as much distance to the UUV and back, reducing effective range. However, laser ranging can provide rich 3D geometric information about the target, including very precise range measurements.
- Passive Tracking Using Ambient Light—In this method only naturally available lighting would be used to image the texture, colour, or shape of the UUV for tracking and pose estimation [7, 8, 9, 10]. No modifications are required to the UUV itself. Previous experiments found that under many conditions of ambient lighting, absorption, and scattering, the quality of target images will not be adequate for feature-based tracking systems. It may be possible to use colour or shape-based algorithms, but their range will be limited under many environmental conditions, and restricted to shallow depths and daylight times.

- Active Lighting—Using a bright light source to illuminate a passive UUV target could possibly improve performance, and also requires no modification to the UUV. However, this approach also suffers from the return path problem, and further results in backscatter from particles in the water that blind the camera system.

Further discussion on underwater optical imaging can be found in [11].

For this project, we have chosen to track LED light sources for position and pose estimation, avoiding many of the difficulties of the other methods listed above. Many other UUV docking optical tracking methods have found target detection at practical ranges of from 10 to 100 m [12, 13, 14, 15, 16] depending on the environment and light source. Various geometries and numbers of lights can be used to extract the 3-D pose of a target from a 2D image[13, 15, 17]. Li et. al. [18] use a pair of stereo cameras to further improve the performance of the system.

There is a long history in robotics of using beacons or target points on a vehicle for pose estimation. This is normally framed in terms of the Perspective N-Point Problem [19], which has several methods for solving. A target with three points visible will produce multiple solutions for pose, while using four or more points should produce unique solutions. Many of the approaches in the literature are for tracking robotic vehicles in the air or on land [20, 21], but the techniques are generally applicable underwater. Bosch discusses the pose problem at length for an underwater system [17], including the use of full camera calibration. However, they do not address tracking LED targets over a wide variety of lighting and turbidity conditions, or a wide variety of ranges from the most distant to immediately in front of the camera.

2.3 LED Light Tracking with An Active Dock

For an underwater optical docking system, the camera, tracking equipment, and final docking control can be placed on either the UUV or on the dock (with LED lights placed on the other). However, an active dock is needed to compensate for limited submarine and UUV transverse manoeuvrability, especially because of their relative motion under waves, so the dock must close with the UUV. This means the dock must know the UUV position. This also means the position sensing equipment on the dock (submarine) can be larger, higher power, and higher quality, while reducing complexity and valuable space claim on the UUV.[1]

In this setup, dock cameras can be oriented looking forward, aft, and outward to locate the UUV anywhere in the semicircular area off to the side of the submarine. The LEDs are low power devices that only need to be turned on during the final few minutes of the docking process, and do not represent a significant electrical or physical burden on the UUV.

Preliminary experiments [3] were conducted to collect a series of representative images under realistic conditions to assess the feasibility of tracking LED lights underwater. This resulted

[1] The dock design is discussed in much greater detail in [1].

in several problems being identified that would not have been apparent in, for example, a swimming pool. The experiments identified the following issues:

1. Directional lights on the UUV (such as small angle LEDs or lasers) were tested and shown to be ineffective unless perfectly aligned with the camera, which is impossible for a UUV subject to motion in water. If not facing the camera, these lights are too dim to see, while if aimed directly at the camera, they overwhelm the sensor's dynamic range. Omnidirectional LED lights, or lights with diffusers on them were found to be helpful for this problem.

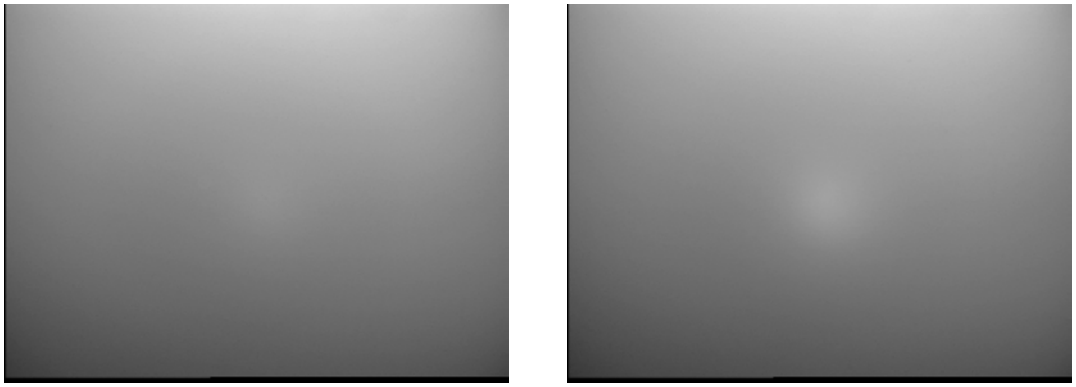
2. Forward Scatter at long ranges in turbid water causes problems in identifying individual lights in an image. In this environment, several light sources tend to appear as a diffuse glowing area, making it difficult to detect and locate the UUV (Figure 3).

3. Ambient lighting and increased turbidity in shallow water can swamp the target lights. Ambient light would be a bigger problem in the open ocean than in the turbid harbour water where these experiments were conducted (Figure 4).

4. Variation in brightness of LEDs in the image over distance due to geometric spreading also posed a problem. The system requires a maximum level of exposure to acquire an LED target at the farthest possible range. However, as the UUV approaches, LED brightness increases and quickly saturates the camera image, making it difficult to distinguish or separate the LED lights (Figure 3).

5. Marine life and other items in the water (such as the UUV hull) can reflect LED or ambient light, and appear brighter than the LEDs themselves (Figure 5 and 6).

In the preliminary experiments, an off-the-shelf camera with an auto-exposure mechanism did not provide the required control or dynamic range to properly isolate the individual lights in our tests due to the difficulties mentioned. As such, we have implemented computer-vision algorithms and a system for computer-vision controlled camera exposure, discussed below.

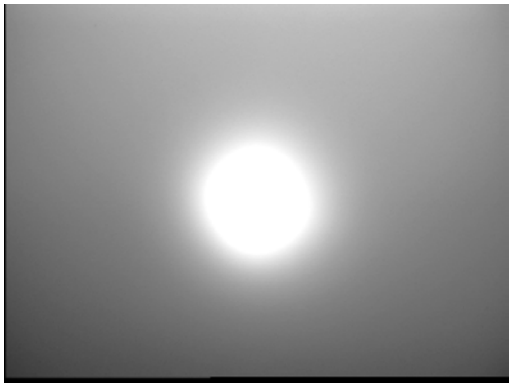


(a) *8m.*

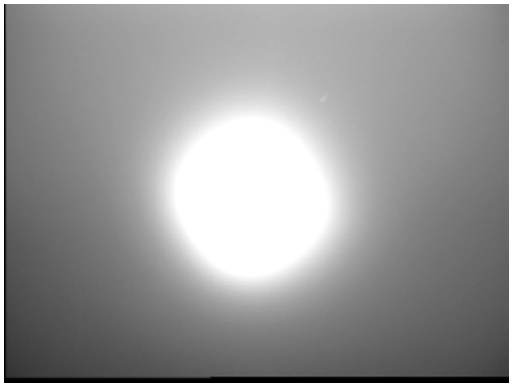
(b) *7m.*

(c) *6m.*

(d) *5m.*



(e) *4m.*



(f) *3m.*

Figure 3: *4 light cluster imaged in turbid water at various ranges.*

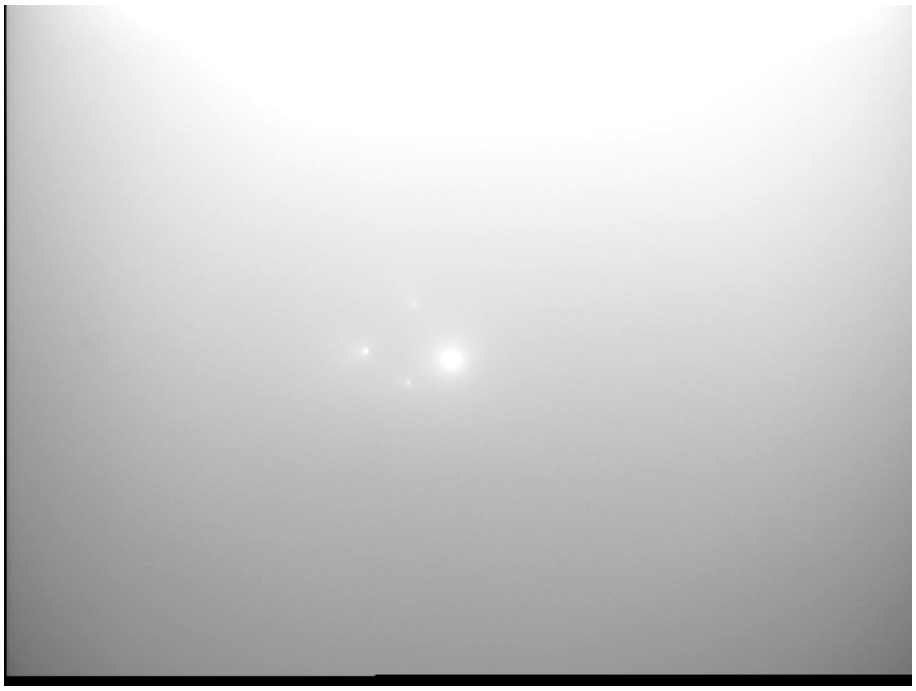


Figure 4: *Bright ambient light at shallow depth making detection of 4 light cluster difficult (centre of image).*

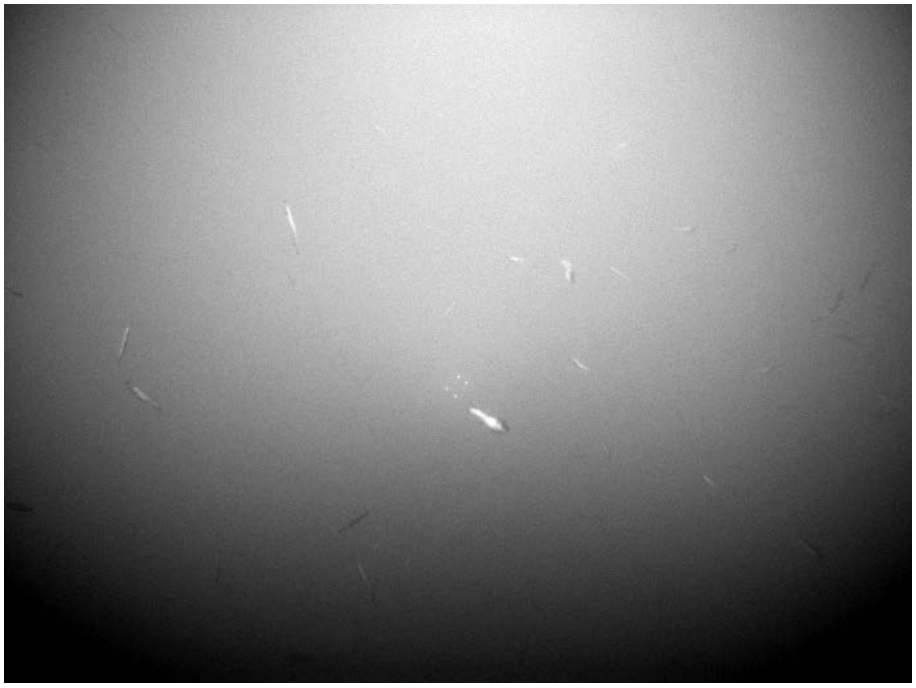


Figure 5: *A school of fish making detection of 4 light cluster difficult (lower centre of image).*

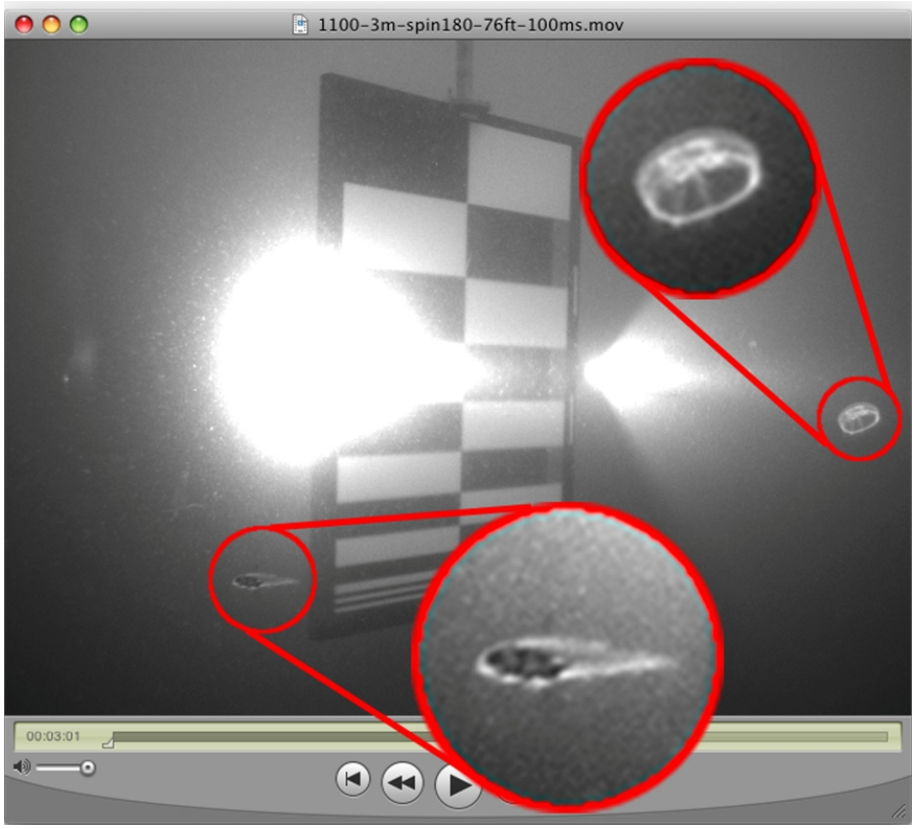


Figure 6: *A particularly difficult image of a black and white test board. Note presence of fish, jellyfish, and large over-exposed light sources.*

3 Computer-Vision Algorithms

The goal of the computer-vision portion of this optical tracking system is to locate the LED targets precisely in the image despite changes in ambient lighting, turbidity, range, etc. From our previous experiments, it was apparent that the effects of the environment must be managed to do this effectively. As such, the computer-vision software contains a number of sub-components:

- A thresholding algorithm that works best for larger, brighter targets.
- A gradient detection algorithm that works best for relatively dim targets.
- A scoring system that selects the most likely location of LED targets from the two image processing methods, depending on whether this system is in the initial stages of acquiring the UUV at longer ranges, or later stages of docking at shorter ranges.
- An exposure control algorithm that iteratively adjusts the camera based on the results of the first two algorithms in order to maintain the LED lights in the image at an appropriate size.

The challenges that these three algorithms must overcome are:

- Dim, diffuse LED light targets due to a combination of relative angle of the light, the amount of ambient light in the water, and absorption and scattering in the water, particularly with range.
- Lights blurring together due to overexposure of the camera sensor or turbidity in the water.
- The apparent centre point of the light not coinciding with the location of the light source due to overexposure or scattering.
- Other bright targets or noise in the image due to ambient light, illumination of marine life, or sensor noise at high exposure levels.

The software was implemented in C++ on a laptop PC, making extensive use of the Open CV libraries [22]. It runs in real time providing position updates at 10Hz or faster on a laptop PC, with processing speed normally being limited by the camera shutter speed (i.e. at most depths the PC executed the software faster than the amount of time the camera needed to gather enough light to image the scene). The pose estimation algorithm was implemented in Maple, as described below.

3.1 Intensity Threshold Detection

In the first method of finding targets, a threshold is used to select the brightest parts of the image as candidates for possible LED locations. The threshold level above which a pixel is

considered to represent the light source is adaptively chosen at each image frame based on the overall image intensity, in order to compensate for changing ambient light conditions.

The algorithm is as follows:

1. Calculate threshold level as a percentile of the overall image intensity. In order to ensure that only the brightest spots in the image are chosen as LED candidate locations, the threshold is set so that only the top 1% to 0.001% of pixels are chosen.[2]

2. Threshold the image, marking everything white that is above this intensity, and everything black that is below this intensity threshold. Figure 7a shows the original image, while Figure 7b shows the thresholded image.

3. Use erode and dilate functions to remove small points of light that are probably noise sources (typically anything smaller than 3 pixels wide). This would theoretically reduce the range of the system in extremely clear water by getting rid of small points of light. However, in general, light sources will be larger than this due to blooming of the light source from scattering of particles in the water, particularly if the exposure is set correctly, as discussed below.

4. Choose predicted LED locations as the centre of mass (centroid) of the remaining continuous white areas in the image (Figure 7c).

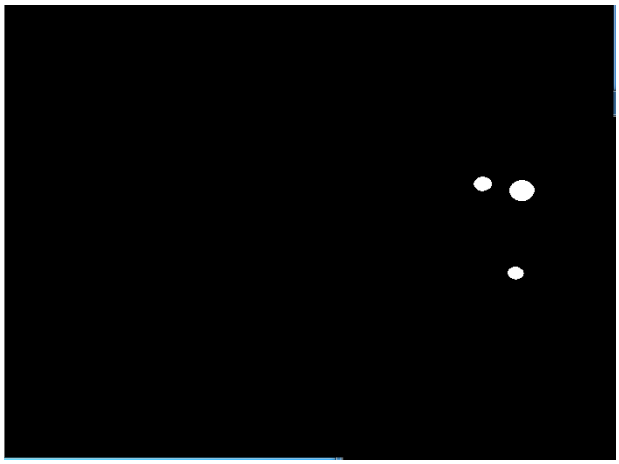
This method works best for bright light sources in clear water, or when the UUV is near the camera and the light sources are large and distinct in the image.

The method of choosing the centroid of the bright point as the predicted LED location might be a poor approximation for lights that are large or have bloomed into odd shapes because of turbidity and odd LED/camera angles (Figure 6). However, with proper exposure control, the LED lights should never become very large in the image, and so the approximation should be a reasonable one.

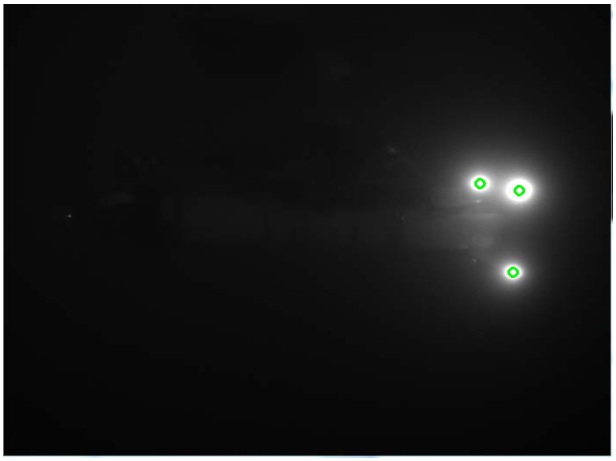
Threshold detection on a set of test lights is shown in Figure 7. This is an image of a mock UUV mounted at 80 degrees to the camera. Three tail LEDs are visible, while the nose and one tail LEDs are not visible. In this example, the algorithm does a reasonably good job of picking the centre of the LED source.

[2] Depending on the stage of docking, the percentile value is changed in real time, as discussed in later sections.

(a) *Original image.*



(b) *Intensity threshold image.*



(c) *Calculated centres (green circles).*

Figure 7: *Intensity threshold detection of relatively bright, large LED exposures.*

3.2 Gradient Detection

To complement the intensity threshold algorithm, the gradient detection method uses the well-known Sobel operator to find the areas of the image with the most change in brightness [23]. The Sobel operator is basically a discrete differentiation operator, meaning that it is approximating the first derivative or rate of change in intensity of the pixels over a specified area. This enables it to find areas of the image that may not be that high in intensity, but for which there is potentially a dim light source causing an increase in brightness.

The algorithm for finding light sources is very similar to the threshold algorithm above, with the addition of a few extra steps at the beginning:

1. Apply the Sobel operator to the image (Figure 8a). We use the sum of the gradient in both the horizontal and vertical directions, equally weighted. When using the Sobel operator, we apply over a kernel size of 5 pixels, which was found to give reasonable results for our environments (Figure 8b).
2. Blur the image to smooth out noise and find areas of high gradient in the image (as opposed to small points of high gradient that are more likely to be noise). We use a normalized box filter for this step.
3. Threshold the image, marking areas as white that are above a certain change in intensity, marking everything else black. Once again, this is done based on choosing the pixels in the image for which the rate of change is higher than the 99th percentile for the entire image (Figure 8c).[3]
4. Use erode and dilate functions to remove small points of intensity change that are probably noise sources (typically anything smaller than 3 pixels wide).
5. Choose predicted LED locations as the centre of mass (centroid) of the remaining continuous white areas in the image (Figure 8d).

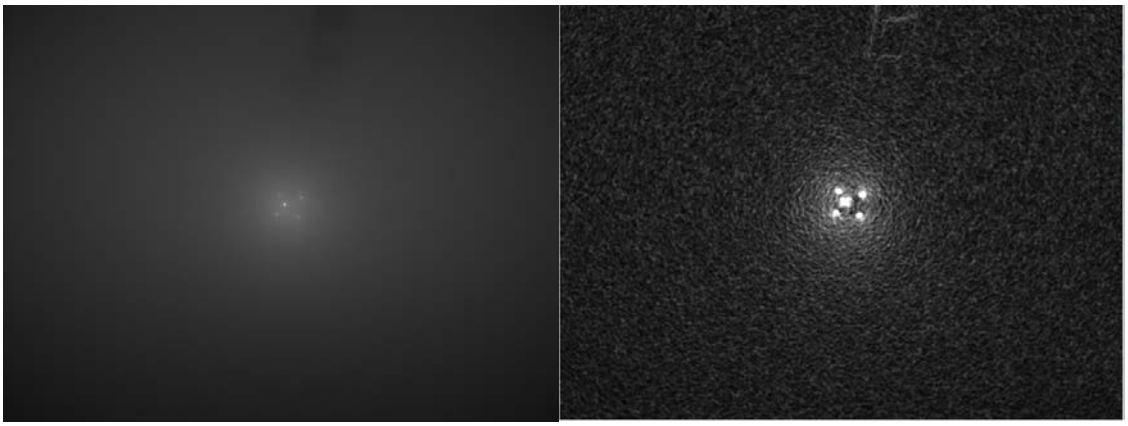
As mentioned, the gradient detector works better than the threshold detector where the light source may appear as a dim or diffuse point of light in the image (Figure 8a). It is also more effective when there is higher ambient light that creates a bright arc of lighter pixels in the top half of the image (Figure 4). In this case, the ambient light creates areas on the top half of the image with higher brightness, but where the rate of change is still low. As such they are rejected as potential light sources.

Small, dim, and diffuse LED sources that are seen at longer ranges and/or higher turbidity have the opposite effect, creating areas that may still be relatively low intensity compared to the overall image, but for which the rate of change is higher in the local area of the image. This makes the gradient detector more effective under these conditions.

One flaw in this method is that for large, bright, LED sources in the image, the gradient detector does not work well to find the centre of the LED. This is because the rate of change

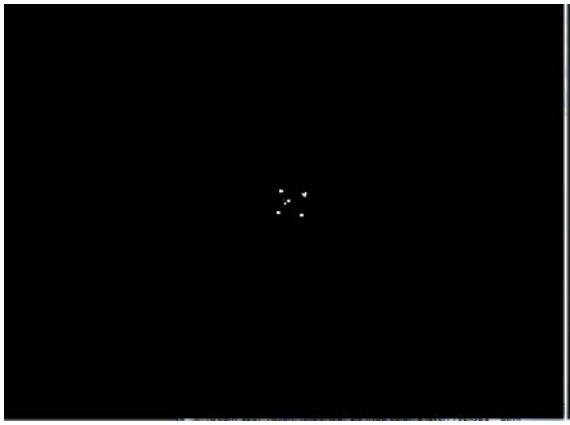
[3] As for the intensity threshold detector, this percentile value is changed as docking progresses.

in intensity is higher at the edges of the light source, than in the middle, which causes this algorithm to pick the edges of large light exposures as a potential LED source. For this reason it is used mostly at the start of docking sequence not the end.

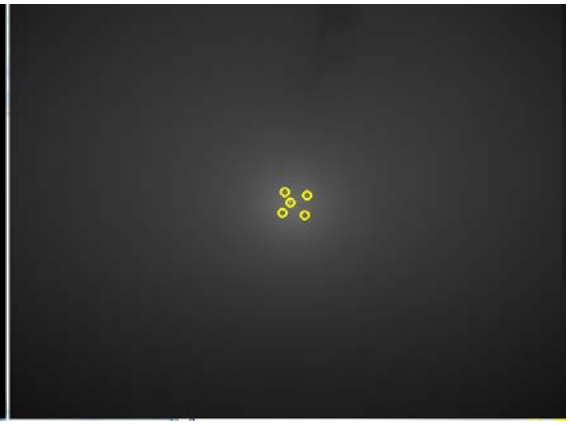


(a) *Original image.*

(b) *Gradient image.*



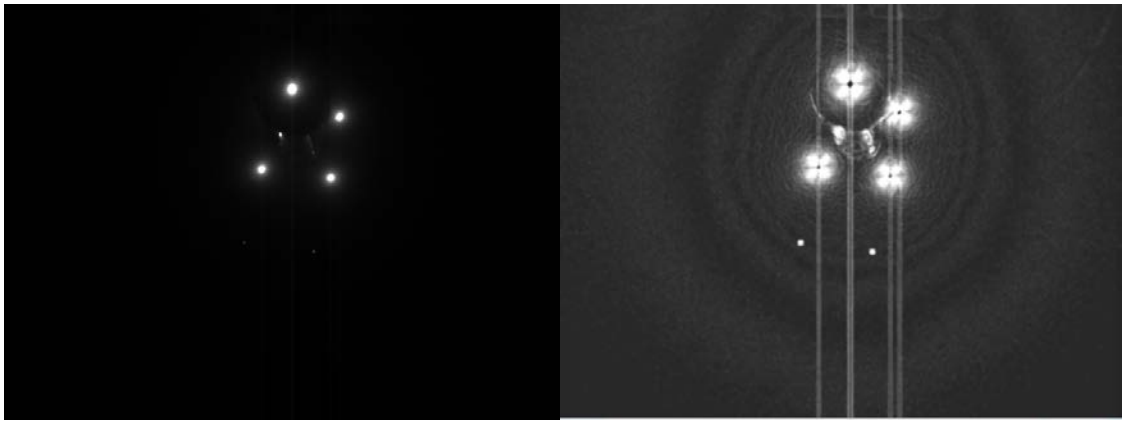
(c) *Gradient threshold image.*



(d) *Calculated centres (in yellow).*

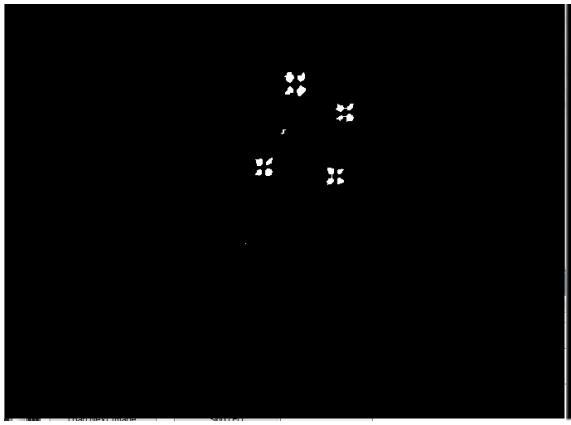
Figure 8: *Gradient detection of relatively dim LED exposures.*

To illustrate the complementary nature of these methods, while the gradient detector worked well for the image in Figure 8, the intensity threshold method for this image was susceptible to noise and objects in the water if tuned to be sensitive enough for these lights. At the same time, for the large bright LEDs in Figure 9 where the threshold detector works well, the gradient method produces a variety of random locations not at the centre of the LED light source, and is not useful in those circumstances (Figure 9d).

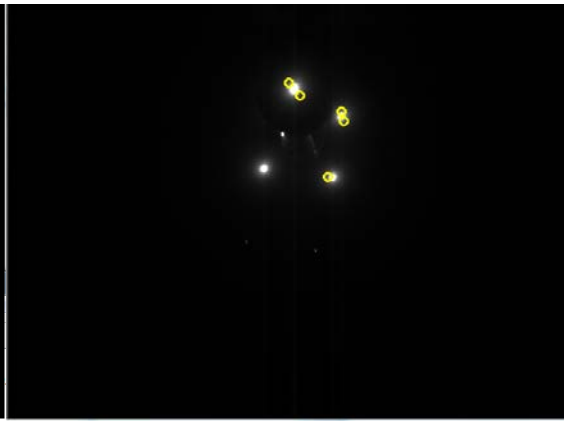


(a) *Original image.*

(b) *Gradient image.*



(c) *Thresholded gradient image.*



(d) *Result using gradient detector only (yellow).*

Figure 9: *Failure of gradient detection of relatively bright, large LED exposures.*

3.3 LED Selection and Tracking

Because of the complementary nature of the threshold and gradient trackers, the results from both are combined in an intelligent way to maximize tracking performance. The first step is to assign a relative certainty value to the candidate LED locations selected by the algorithm above. The certainty value is based on the LED location's intensity and size because larger, brighter targets are more likely to be LED sources than noise or reflections off of the UUV or marine life.

For LED sources found by the threshold algorithm, the certainty is calculated as follows:

- For very bright areas of at least 10 pixels in area, we assign certainty of 1.0, as extremely bright areas above a minimum size are most likely LED sources. Very bright is designated as having intensity of at least 90% of maximum intensity possible (i.e. for 8 bit images this is having and intensity value of 230 out of 255).
- For bright areas of at least 20 pixels in area, we also assign a certainty of 1.0, as large, bright areas are also most likely LED sources. Bright is designated as having intensity of at least 70% of maximum intensity possible.
- For large areas that are less bright ($>$ 20 pixels, intensity less than 70% of maximum), a certainty is assigned as the intensity percent of maximum (i.e. area intensity/maximum intensity).
- For small areas that are less bright ($<$ 20 pixels, intensity less than 70% of maximum), a certainty is assigned as the intensity percent of maximum, multiplied by the fraction area/20 pixels. (i.e. (area/20pixels)*(intensity/maximum intensity)).

The certainty for candidate areas detected by the gradient method is assigned as the rate of change in intensity divided by the maximum possible rate of change. This generally results in certainty values that are less than those for the threshold method. In addition, its certainty is multiplied by a scaling factor to reduced emphasis on this tracker later in later docking phases.

The threshold method typically provides more reliable results when the LED light source is obvious, with a better estimate of the LED position in the brighter part of the image. In situations where there are few obvious LED sources in the image, then the gradient method will provide greater sensitivity at the cost of the potential for false positive results, and less accurate estimation of LED position.

The last step is to combine the results of the threshold and gradient methods. The candidates are first sorted to remove any that are too close to each other (and therefore probably the same LED). Then, they are sorted from highest to lowest certainty value.

3.4 Tracking and Exposure Control

The docking scenario will have the UUV moving over the full range of the camera system, from beyond the visible range, to immediately in front of the camera. As such, the LED's appearance in the image varies drastically as well (Figures 10 and 11). It was found that exposure control of the camera is critical to provide any level of functionality over the full range of docking, and that commercial auto-exposure systems were not appropriate for the image processing tasks. Exposure control also provides robustness to variable ambient light conditions due to depth, turbidity, etc.

(a) *10 ms* ***(b)*** *100 ms* ***(c)*** *1000 ms*

Figure 10: *Exposures of a 5 LED cluster at 14m range.*

(a) *10 µs* ***(b)*** *100 µs* ***(c)*** *1 ms* ***(d)*** *10 ms* ***(e)*** *100 ms*

Figure 11: *Exposures of a 5 LED cluster at 2m range (top two LEDs obscured by UUV body).*

In order to provide dynamic range to the computer-vision system, two tactics were used:

- Controlling the camera exposure during the docking process using "states" of LED tracking: *Initializing*, *Acquiring*, *Tracking*, and *Lost*. This allows us to have maximum sensitivity when trying to first find the UUV lights at long ranges, while still accurately finding the LED positions in the image at shorter range.
- Increasing how selectively the system accepts results from the computer-vision algorithm based on what "state" it is in. This also allows the system to have maximum sensitivity at long ranges, and reduces false positives and errors in LED positions at shorter ranges.

Below is a description of the system functionality in each state:

1. *Initializing*—At system start-up, it is assumed that the UUV and LED lights are not visible in the image (i.e. the UUV will be at ranges far greater than that of the camera). As such, the camera exposure is started very bright, and decreased until no

targets are identified by the computer-vision algorithms, in order to reject noise from ambient light or marine life. This could potentially reduce the range of the system, but provides us the maximum sensitivity without incorrectly identifying the target. This exposure is held for several seconds to ensure that this level is stable before moving on to the next state.

2. *Acquiring*—The system now waits for the UUV LEDs to become visible. In this state, the exposure is held constant, waiting for results from the threshold and gradient detectors when the LEDs appear. At this point, the certainty that the system is willing to accept in results from the algorithms for candidate LED areas is greatly lowered, so that our sensitivity and range is maximized. Under most conditions, it is most likely the gradient detector that will first find the LED source in the image during this stage. Once an LED target is first found, we wait to see it in the same spot for several seconds, to ensure noise rejection and stability before reporting LED locations. At this point, the system moves on to the next state.

3. *Tracking*—In this state, the system begins to report the location of LEDs in the image for pose estimation. Once an LED source is acquired, the certainty that we require from candidate light sources is increased, to improve noise rejection from reflections off the UUV hull, marine life, etc., while still maintaining tracking of our UUV LEDs. In this state, the system will continually acquire more LED sources (applying higher certainty requirements), until it has the number of LEDs that are mounted on the UUV. Sources that move in the image too far and too fast are also rejected. In this state, it also applies exposure control to the camera, decreasing the exposure as the LED sources become too bright and large in the image, and increasing the exposure if they start becoming too small and dim. This keeps them at the ideal size in the image to accurately pinpoint the location of the LED light for pose estimation.

4. *Lost*—If the LEDs are lost during the tracking phase, the camera exposure is gradually increased within acceptable limits in an attempt to re-acquire the LEDs. However, if a certain time elapses before the LEDs are found, the system resets to the *Initializing* state.

When adjusting the exposure level in the "Tracking" state, a simple step method was used to either increase or decrease the exposure time by a certain percent whenever the target was the wrong size (typically 20% longer or shorter exposure time). The limits were not so critical, but a pixel area for the largest threshold target less than 40 pixels would cause an increase in exposure, while an area greater than 70 pixels would cause an increase in exposure. Although this method is fairly simplistic, it proved robust for limited underwater dynamics conditions, and did not result in continuous "hunting" for the correct exposure level.

During the Acquiring and Tracking phases, there are a number of parameters that should be adjusted on-the-fly to maintain good location-finding of the LEDs. Among these are:

- The intensity percentile number for the threshold and gradient trackers, which is the minimum intensity of the pixels for which we will accept them to be from an LED source.
- The number of pixels of separation between candidate LED sources for which we will consider them to be from the same source or different sources. This is particularly important when the two trackers are both active, as the threshold tracker might find the centre of light, while the gradient tracker might find the edge of a light. In this case we would want to categorize them as the same light source even though they might be in slightly different image locations.
- The minimum certainty value that each detection must have to be considered an LED light source.
- The certainty weighting that we give to sources from the gradient tracker vs. the threshold tracker.
- The amount we erode and dilate the threshold locations to remove noise.
- The minimum size (in pixels) of target that we will deem to be a valid LED location.

At the farthest ranges, during the start of acquisition, the LEDs will appear the dimmest and closest together. During this phase, we will adjust the percentile number and the minimum size to be lower, such that dimmer LED sources will be considered LED sources (without setting it so low that we pick up noise). Also, as discussed, the minimum certainty required for an LED source detection is lowered, and the amount we erode and dilate the image is lessened so that we don't erode away light sources. Finally, in the initial tracking phases we give the gradient tracker higher certainty values than the threshold tracker.

As we continue with tracking the LEDs, and the UUV gets closer, the LEDs get bigger, brighter, and farther apart (depending on pose). In addition, at the closest ranges we may start getting reflections off the UUV hull that may be confused with light sources. As such, we must make a number of adjustments to maintain tracking as we progress. The first is that we must reduce the reliance on the gradient tracker algorithm, giving preference to threshold detections by increasing those certainty values. We must also increase amount that we erode and dilate the threshold images, and the minimum certainty value and minimum size that we will accept as a valid detections, in order to provide better noise rejection.

Finally, due to the fact that as the UUV nears, our separation between the LEDs should increase, we must adjust the minimum amount of pixel separation between candidate LED sources allowed to improve noise rejection. This has one flaw in that a UUV posed at a high angle to the camera may have LEDs that appear close together or may be blended together, which provides for a difficult imaging problem. However, we must make the assumption that by late stages in docking, the UUV will be relatively well aligned, and this will be a minimal problem. An example of this is shown in Figure 12, where the UUV is almost 90 degrees to the camera. The UUV nose LED is on the left, while 3 tail lights are visible on the right. The two top LEDs have almost merged into one.

Figure 12: *Difficulty discerning LEDs due to a high relative orientation of UUV.*

3.5 Position and Pose Estimation

Once the LEDs have been located in the image, the next task is to find the UUV relative position and pose (orientation) to the dock mechanism. If there is only one LED or point of light visible, and only one camera looking at it, the system is able to report the horizontal and vertical angles to the UUV target, but not range or any other information. This is still very useful as keeping the light centred in the image as the light approaches the camera provides terminal guidance which has been shown to be effective [12]. In terminal guidance mode, absolute positional accuracy is less important than dock dexterity, reach, and control. Terminal guidance is expected to be used after the initial sighting of the UUV while it is still too far away to discern individual lights, and possibly for final docking if a camera can be located near the capture point on the dock.

Once individual LEDs can be identified, or if multiple cameras are used, the system can report position and pose using the coordinates of the LEDs in the camera images together with their known locations on the vehicle. UUV position is determined by three unknowns and pose by another three. So six pieces of independent information are required to determine position and pose. A single LED in a single camera image provides two pieces of information, namely the pixel coordinates in the 2D image. Therefore, three LEDs in a single image, one LED seen by three well spaced cameras, or two LEDs seen by one camera and one LED by another are the minimum requirements for knowing position and pose. This information enables the most predictive, responsive, and robust docking control system. Our challenge is to make this information available at reasonable update rates.

We consider two cases in this report, though only the first has been implemented, verified, and validated. The first case is when one camera views n LEDs on the UUV, where $n \geq 3$. An experimental validation of this case is presented in the next section. The second case is probably more realistic in an actual docking scenario: it allows for multiple cameras viewing multiple LEDs on the UUV, and accounts for arbitrary camera locations and orientations with respect to a common frame of reference used by the cameras and dock. A theoretical analysis of this latter case, based on the following subsection and associated details in Annex A, is presented in Annex B.

3.5.1 One Camera, Multiple LEDs

The UUV location and pose relative to a single camera are estimated using the setup shown in Figure 13. A right handed Cartesian coordinate system ξ, μ, ν is fixed to the camera with its origin at the middle of the camera charge-coupled device (CCD). The ξ axis is normal to the CCD and points in the direction the camera faces. The μ and ν axes are aligned with the horizontal and vertical pixels, respectively, in the camera image. A camera spherical coordinate system is also useful and gives the range ρ, bearing β, and elevation δ to the UUV:

$$\xi = \rho \cos\beta \cos\delta, \quad \mu = \rho \sin\beta \cos\delta, \quad \nu = \rho \sin\delta \tag{1}$$

The x, y, z axes in Figure 13 are standard body-fixed axes for underwater vehicles [24, 25].

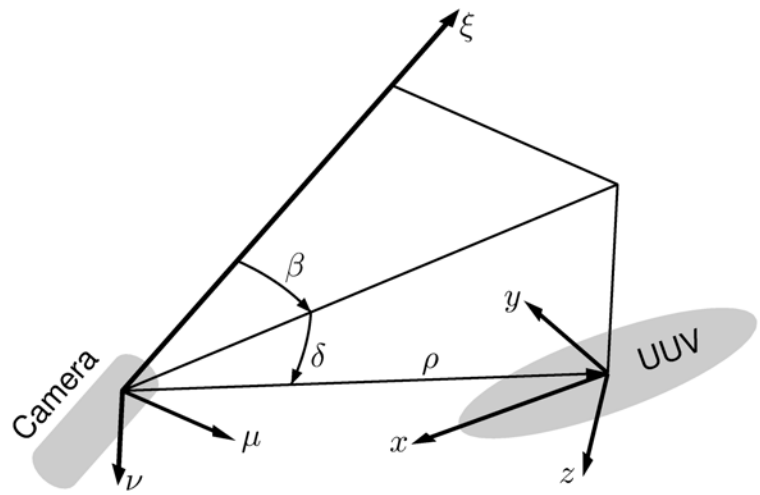


Figure 13: *Camera and UUV coordinate systems.*

Their origin is typically on the UUV hull centreline at or close to the vehicle centre of buoyancy; the x axis points forward along the hull centreline, the y axis points to starboard, and the z axis points downwards through the keel. The UUV pose, relative to the camera, is defined using a standard underwater vehicle Euler transformation. If the camera and UUV axes are initially aligned, UUV pose is achieved by first yawing an angle ψ about the z axis, then pitching an angle θ about the y axis, and finally rolling an angle ϕ about the x axis. Thus, UUV body axes vectors $\mathbf{X}$ are reoriented to camera axes using the Euler transformation $\mathbf{A} \cdot \mathbf{X}$ where:

$$\begin{aligned}\mathbf{A} &= \begin{bmatrix} \cos\theta\cos\psi & \sin\phi\sin\theta\cos\psi - \cos\phi\sin\psi & \cos\phi\sin\theta\cos\psi + \sin\phi\sin\psi \\ \cos\theta\sin\psi & \sin\phi\sin\theta\sin\psi + \cos\phi\cos\psi & \cos\phi\sin\theta\sin\psi - \sin\phi\cos\psi \\ -\sin\theta & \sin\phi\cos\theta & \cos\phi\cos\theta \end{bmatrix} \\ &= \mathbf{A}(\phi,\theta,\psi) = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.\end{aligned} \tag{2}$$

Let p, q be the pixel coordinates giving location in a camera image. Their common origin is the top left corner of the image and they increase in the same directions as μ, ν. If p_0, q_0 locate the centr of the image, p_j, q_j locate LED j in the image, as shown in Annex C, the variable C is the calibration constant for the underwater camera and housing in seawater in pixels, then:

$$\begin{aligned} P_j &\equiv \frac{p_j - p_0}{C} = \frac{\mu_j}{\xi_j} \quad \Longrightarrow \quad \xi_j P_j = \mu_j \\ Q_j &\equiv \frac{q_j - q_0}{C} = \frac{\nu_j}{\xi_j} \quad \Longrightarrow \quad \xi_j Q_j = \nu_j \end{aligned} \tag{3}$$

where:

$$\begin{bmatrix} \xi_j \\ \mu_j \\ \nu_j \end{bmatrix} = \begin{bmatrix} \xi \\ \mu \\ \nu \end{bmatrix} + \mathbf{A} \cdot \begin{bmatrix} x_j \\ y_j \\ z_j \end{bmatrix}. \tag{4}$$

Here, ξ_j is $x + x'$ from Figure C.1 and μ_j and ν_j are y, depending on whether the horizontal or vertical field of view is being analyzed. The fixed coordinates x_j, y_j, z_j of each LED in UUV axes must be known. For $j = 1, 2, 3$, (3) provides the 6 equations required to solve for the 6 unknowns $\xi, \mu, \nu, \phi, \theta, \psi$ giving the position and pose of the UUV.

The UUV position coordinates ξ, μ, ν appear linearly in (3), by way of (4) and can be eliminated using 3 of the equations. This leaves three equations that are nonlinear in the Euler angles ϕ, θ, ψ and which must be solved iteratively. Newton's method is ideal for this because the formulation (2), (3), and (4) is simple enough that analytic expressions for the necessary derivatives are available and can be rapidly evaluated. CPU time is not an issue. The problem, however, is that there can be several unique solutions and in some situations these can be very close together. A robust pose prediction requires more than three LEDs and these should not be coplanar.

With n LEDs in an image, there are $N = n!/(3!(n-3)!)$ groups of three independent equations to solve for only three unknowns. There will not be a unique solution because of inherent error in the pixel coordinates, so a least squares approach is used. For large n, this is more efficient than solving all the equation groups separately and averaging the results.

In the least squares formulation, we take advantage of the fact that the ξ_j, μ_j, ν_j in (3) are, by way of (4), linear in the a_{ij} from (2). If:

$$\mathbf{V}^T \equiv [a_{11}, a_{12}, a_{13}, a_{21}, a_{22}, a_{23}, a_{31}, a_{32}, a_{33}] \tag{5}$$

then the equations to solve can be put in the form $\mathbf{B} \cdot \mathbf{V} = 0$ where $\mathbf{B}$ is a $3N \times 9$ coefficient matrix (see Annex A) independent of the unknown Euler angles. The sum of the squares S of these equations is:

$$S = (\mathbf{B} \cdot \mathbf{V})^T \cdot \mathbf{B} \cdot \mathbf{V} = \mathbf{V}^T \cdot \mathbf{B}^T \cdot \mathbf{B} \cdot \mathbf{V}. \tag{6}$$

The nice result here is that the inner product $\mathbf{B}^T\mathbf{B}$, which has constant coefficients, is reduced to a 9×9 symmetric matrix that only needs to be calculated once; using iteration to solve for the pose uses $\mathbf{B}^T\mathbf{B}$, not $\mathbf{B}$.

Newton's method is again used to minimize S. This requires the gradient $\mathbf{G}$, the iteration correction $\mathbf{\Delta}$, and the Hessian $\mathbf{H}$:

$$\mathbf{G} = \begin{bmatrix} \partial S/\partial\phi \\ \partial S/\partial\theta \\ \partial S/\partial\psi \end{bmatrix}, \quad \mathbf{\Delta} = \begin{bmatrix} \Delta\phi \\ \Delta\theta \\ \Delta\psi \end{bmatrix}, \tag{7}$$

$$\mathbf{H} = \begin{bmatrix} \partial^2 S/\partial\phi^2 & \partial^2 S/\partial\phi\partial\theta & \partial^2 S/\partial\phi\partial\psi \\ \partial^2 S/\partial\theta\partial\phi & \partial^2 S/\partial\theta^2 & \partial^2 S/\partial\theta\partial\psi \\ \partial^2 S/\partial\psi\partial\phi & \partial^2 S/\partial\psi\partial\theta & \partial S^2/\partial\psi^2 \end{bmatrix}.$$

which are derived in Annex A.

S is minimized when $\mathbf{G} = \mathbf{0}$ (3 equations in 3 unknowns) but to ensure the solution is not a saddlepoint or maximum, $\mathbf{H}$ must be positive definite. A pure Newton approach makes an initial guess for the pose and then repeatedly solves:

$$\mathbf{G} + \mathbf{H} \cdot \mathbf{\Delta} = \mathbf{0} \tag{8}$$

for a series of corrections $\mathbf{\Delta}$. This method was unreliable for processing many of the 3 LED images. Therefore a modified Newton method called a "restricted step" method by Fletcher [26] and a "trust-region" method by Nocedal and Wright [27] was adopted. The method limits $|\mathbf{\Delta}|$ to some value Δ_{max}. The pure Newton method is allowed to proceed normally unless $|\mathbf{\Delta}|$ exceeds Δ_{max}, S increases, or the Hessian is not positive definite. Δ_{max} is adjusted up or down depending on the algorithm's ability to decrease S. If necessary, $\mathbf{H}$ is made positive definite by increasing the magnitude of its diagonal elements by $\lambda \geq 0$ following:

$$(\mathbf{H} + \lambda\mathbf{I}) \cdot \mathbf{\Delta} = -\mathbf{G} \quad \text{where} \quad \lambda(\Delta_{max} - |\mathbf{\Delta}|) = \mathbf{0}. \tag{9}$$

The Newton and modified Newton methods have been customized for a 3×3 Hessian with $\mathbf{G}$ and $\mathbf{H}$ calculated analytically. The images discussed below were processed by running Maple scripts in hardware floating point mode; the CPU time to process each image averaged 0.03 s. The initial condition was the solution from the previously processed image. A total of 909 validation images were processed and all were processed successfully. Of these, 682 contained 4 or 5 LEDs and the remainder only 3 LEDs. None of the former group required use of the modified Newton method but 50 images from the latter group did. The pure Newton method solutions of (8) averaged 3.3 iterations per solution to reduce $|\mathbf{\Delta}|$ below 10^{-4} radians, whereas the modified Newton method averaged 5.4 iterations per solution. The maximum number of iterations required for a single solution was 9.

A prior version of this analysis [28] which did not sort each group of three LEDs, as described in Annex A, and just used the best conditioned three LED group to determine ξ, μ, ν, had to use the modified Newton method 105 times on the same dataset. These calculations averaged about 12 iterations per solution and one was aborted when the iteration count exceeded 100. Furthermore, for all 909 calculations, the total iteration count was 20% larger and the total processing time 15% longer than for the method described in Annex A.

Despite the improved processing robustness provided by the Annex A conditioning, only minor or no improvements were obtained in overall validation errors discussed in Section 5.2 (Table 1). The biggest improvement was in the range prediction where the standard deviation in the error reduced from 1.4 to 1.3 percent of range (for 4 or more LEDs).

4 Experimental Setup

In order to test the effectiveness of the computer-vision and pose estimation systems, experiments were conducted at DRDC – Atlantic Research Centre in October 2016. This section describes the hardware and facilities used to conduct these tests.

4.1 Underwater Video Cameras

The camera used in these experiments was a low-light SubC uLux, custom built for this project and since commercialized as the StarGazer. It is based on the Retiga 1350B (QIClick) scientific video camera from QImaging, with 1392 x 1040 resolution[4], and minimum illumination of 0.000001 Lux. It has software-controllable gain and exposure settings, and was used with a fixed-focus lens giving a horizontal field of view of about 40 degrees in water. IEEE-1394 digital video from the camera is relayed with a pair of FireNEX-COAX-S800 IEEE-1394 repeaters and through 25 metres of coaxial underwater cable. The images were captured from the IEEE-1394 interface using a laptop computer.

As a backup, all data was also captured with less expensive Shark Marine SV-DSP-ZOOM2 camera. It has a computer-controlled variable zoom lens with a field of view of about 46 degrees, as well as software controllable exposure. Minimum illumination is 0.1 lux. The camera was controlled via a CANBUS protocol from the laptop computer. NTSC video was also captured on the laptop using an Imperx framegrabber at 640x480 (interlaced) resolution.

(a) *SubC uLux Scientific camera* ***(b)*** *Shark Marine SV-DSP-ZOOM2 camera*

Figure 14: *Cameras used to gather underwater images.*

In analyzing our data, it was found that for the LED powers and distances we experimented with, the 0.1 lux Shark camera was more than adequate to gather enough light to image

[4] Higher resolution may result in slightly higher accuracy at the price of execution time. In these tests, the camera was used at a resolution of 696x520.

the LED sources. However, the analog transmission of video and interlaced images added significant noise to the images that makes the computer-vision aspects much more difficult (Figure 15). As such, it is recommended that cameras for use in this application use the highest resolution available and digital transmission methods.

(a) *SubC uLux sample image.* ***(b)*** *SV-DSP-ZOOM2 sample image.*

Figure 15: *Sample images.*

4.2 LED Lights and UUV Surrogate

In order to provide configurable and adjustable LED light sources on the UUV, a custom cable was made containing 6 white Cree XLamp XP-E LEDs (Figures 16 and 18). These LEDs are capable of producing more than 250 lumens of light each at 1A of current. During our experiments, it was found that this brightness was excessive for the ranges available to us in the experimental setup. As such, they were generally operated at much lower current than this to reduce the light entering the camera to be within its dynamic range capabilities (i.e. to avoid washing out the image).

These LEDs have a wide viewing angle of 115 to 130 degrees, and transmitted light symmetrically over this angle. In contrast to LED light sources used in our previous experiments, diffusers were not required, as the brightness in the image was similar enough between straight-on viewing and off-angle viewing. The LEDs were sealed in a Hysol RE2038 and HD3404 compound, with the LED kept as close as possible to the potting surface for viewing angle and heat dissipation. A separate control box was used to allow switching of the LEDs on or off for experimental purposes.

To simulate the proportions of a UUV, the LEDs were mounted on a surrogate made from PVC pipe to approximate an Iver2 vehicle. One LED was mounted on the nose, and one on the end of each tail fin (Figures 17, 18 and 19). This configuration provides an adequate tradeoff between having several LEDs visible at all poses, with having maximum separation between the lights, and without lights overlapping and interfering with each other.

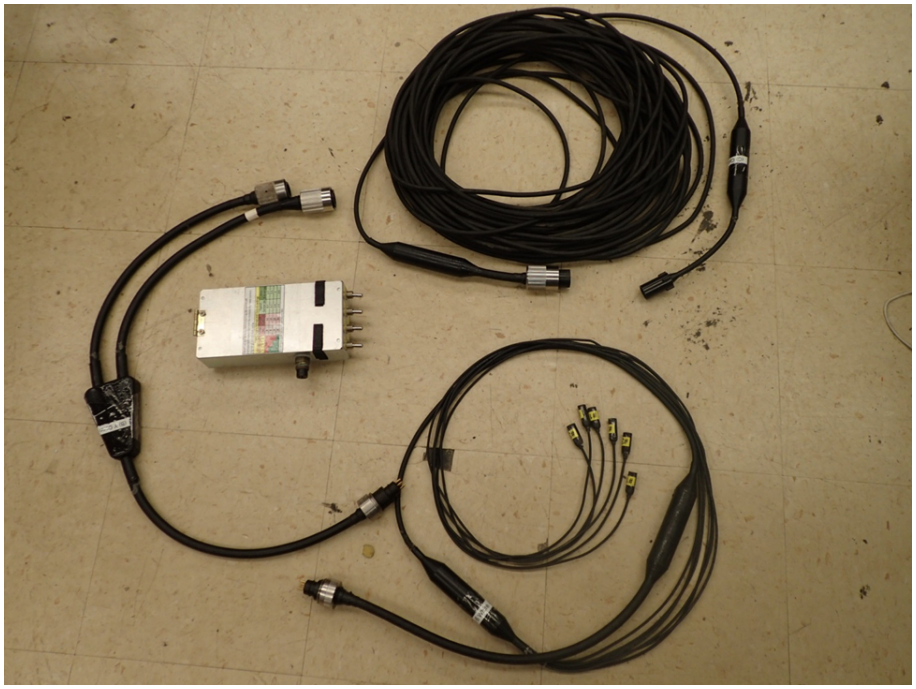


Figure 16: *LED light control box and cables.*

A single LED on the nose also provides the ability to do pure "homing", where the dock could ignore all of the pose information of the UUV, and keep the nose LED in the centre of the image for terminal docking.

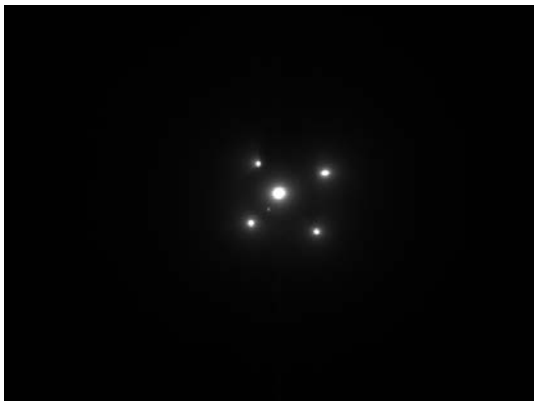


Figure 17: *LED configuration as seen underwater (from the nose).*

(a) Potted LEDs.

(b) LED Closeup.

Figure 18: *LED light sources used for underwater testing.*

Figure 19: *UUV surrogate used for experimentation.*

4.3 Barge Facility

Tests were conducted at the DRDC acoustic calibration barge facility in Halifax's Bedford Basin. Shown in Figure 20, the barge is held in place by four permanent mooring anchors and is sited at a location with a water depth of 42 metres.

The barge provides a sheltered work space with unfettered access to the ocean beneath, and approximately half of the interior floor area is a "well" that is open to the ocean water. The interior, shown in Figure 21, is focused on the 18 m by 9 m well. Two movable bridges and platforms (referred to as cross trolleys), rotating stations, and an overhead crane allow heavy equipment to be suspended at specific depths and locations beneath the barge. A close-up picture of a well bridge is shown in Figure 21b.

Figure 20: *Exterior view of DRDC – Atlantic Research Centre's Acoustic Calibration Barge in Bedford Basin.*

In operation, electric motors allow the well bridges to move along the length of the well, and each cross trolley to be moved transversely across the well on the bridge. At the same time, the rotation stations are used to lower a series of metal shafts (hangers), to control depth of the test equipment (Figure 22), giving full x, y, z positioning of test equipment in the well. Further, each of the rotation stations can rotate freely to spin the device at the bottom of their shafts to any horizontal angle.

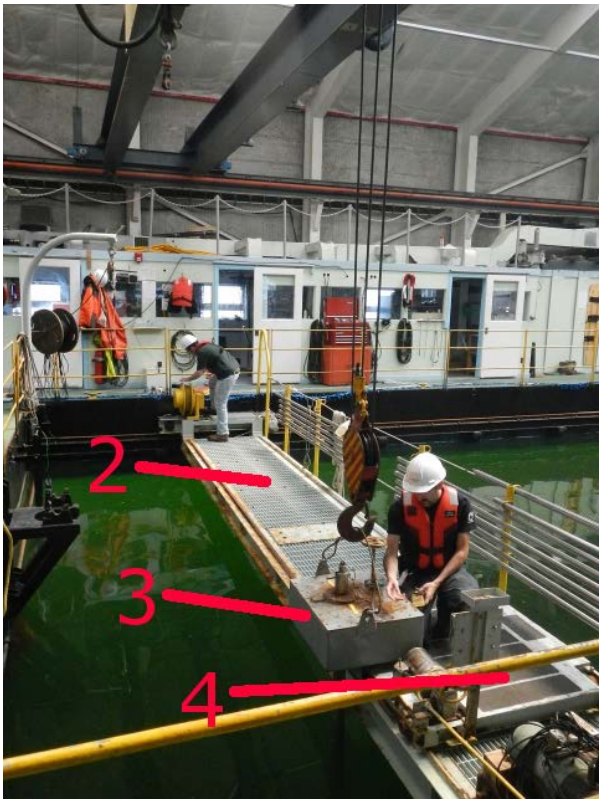
In these tests, the UUV surrogate was suspended at the bottom of the stern station (below the stern cross trolley and bridge), while the camera equipment was suspended on the bow station (below the bow cross trolley and bridge). In this way, a number of relative poses

can be established between the camera and surrogate UUV. The only degrees of freedom on this setup that were not controllable were UUV pitch and roll.

A diagram of the general arrangement for the barge experiment is shown in Figure 23. The moveable optical target UUV is at the stern end of the well (left end), while the cameras, which were kept stationary for the experiments, are at the forward end of the well (right, bow).

One difficulty with this setup is that the equipment is suspended at the bottom of the series of linked hanger shafts below the rotation stations. To prevent the hangers from bending in the current, they were linked by hinged pivot joints. However, this also meant that the equipment at the bottom would drift as if suspended in the water by a rope, leading to uncertainty in the relative position of the camera and UUV surrogate, a problem addressed by the mechanism described in the next section.

(a) *Barge well.*



(b) *Barge well bridge.*

Figure 21: *Interior view of the acoustic calibration barge, showing (1) the internal well, (2) well bridge, (3) rotating station and (4) cross trolley.*

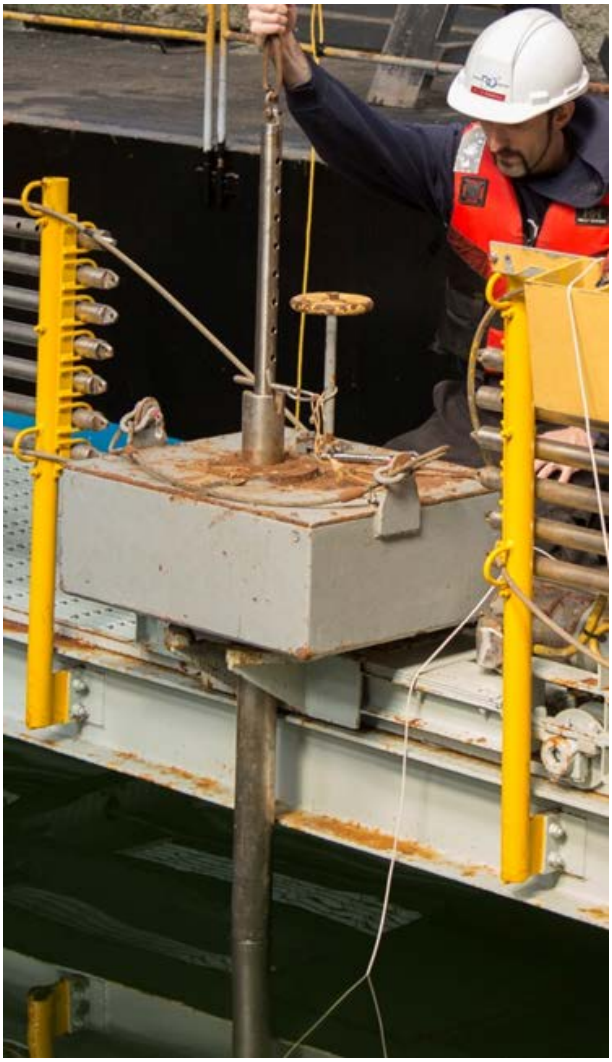


Figure 22: *A view of one of the rotation stations, while lowering a hanger shaft.*

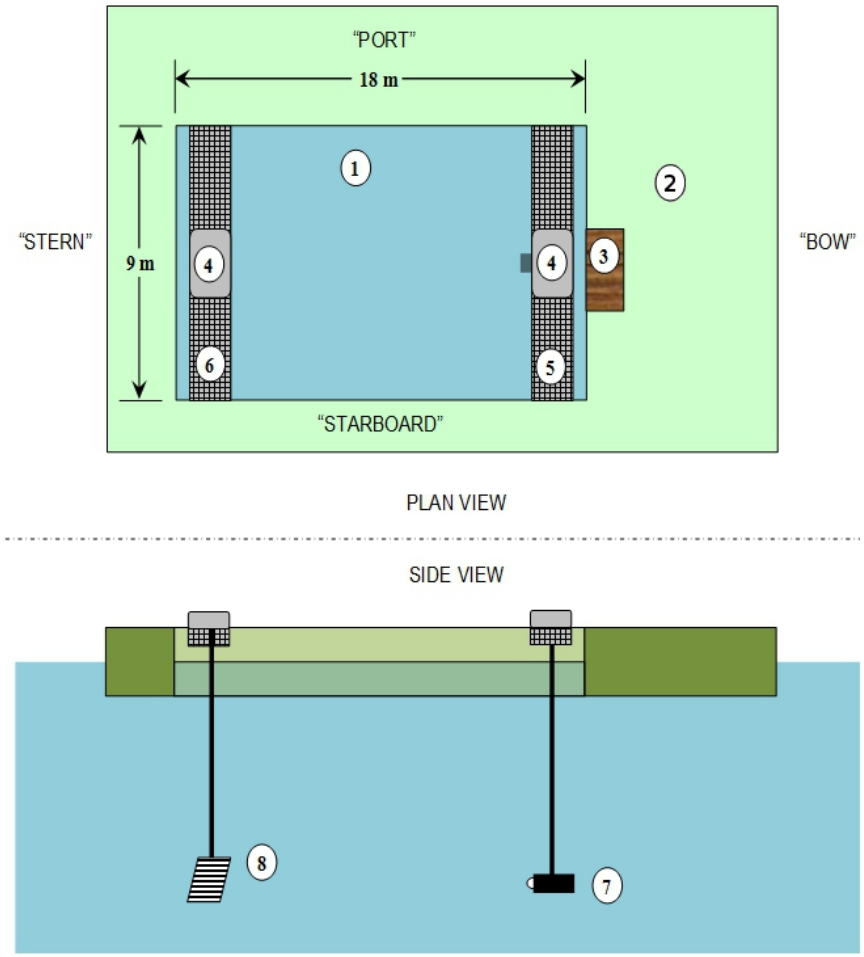


Figure 23: *General arrangement of barge showing (1) well, (2) working deck, (3) work desk, (4) cross-trolleys with rotating stations, (5) bow well bridge, (6) stern well bridge, (7) video camera, (8) UUV surrogate target.*

4.4 Ground Truth Measurements

Much effort was put towards designing and installing a "ground truth" system to position the UUV and cameras in an established relative pose, so that the measurements from the computer-vision system could be verified for accuracy and precision. Previously, a short-baseline acoustical system was used in an attempt to provide this functionality, but was found to be insufficiently accurate. As such, a mechanical linkage was designed to attempt to "fix" the relative positions of the equipment at the bottom of the shafts in the water. This "Pose Validation Rig" (Figure 24) consisted of several components:

1. A longitudinal "range" pole that connects the bottom of the stern and bow vertical hanger shafts, maintaining the camera and UUV equipment at set, pre-measured spacings apart (indexed every 1.0m).

2. Trays that sit above the equipment at the bottom of both the stern and bow vertical hanger shafts. These trays hold weights to stabilize the vertical hanger shafts from swaying in the current, and provide a connection point for the longitudinal index pole to attach to the hanger shafts.

3. Guide cables spooled out from cable pulleys that attach to the trays, allowing the trays and range pole to be raised and lowered, while also providing lateral stability to the hanger shafts and trays.

A full description of the Pose Validation Rig can be found in a previous report [29]. Pictures of the stern and bow station equipment and trays are shown in Figure 25. The installed range pole is shown in Figure 26a. The installed system and cable pulleys can be seen at the surface in Figure 26b.

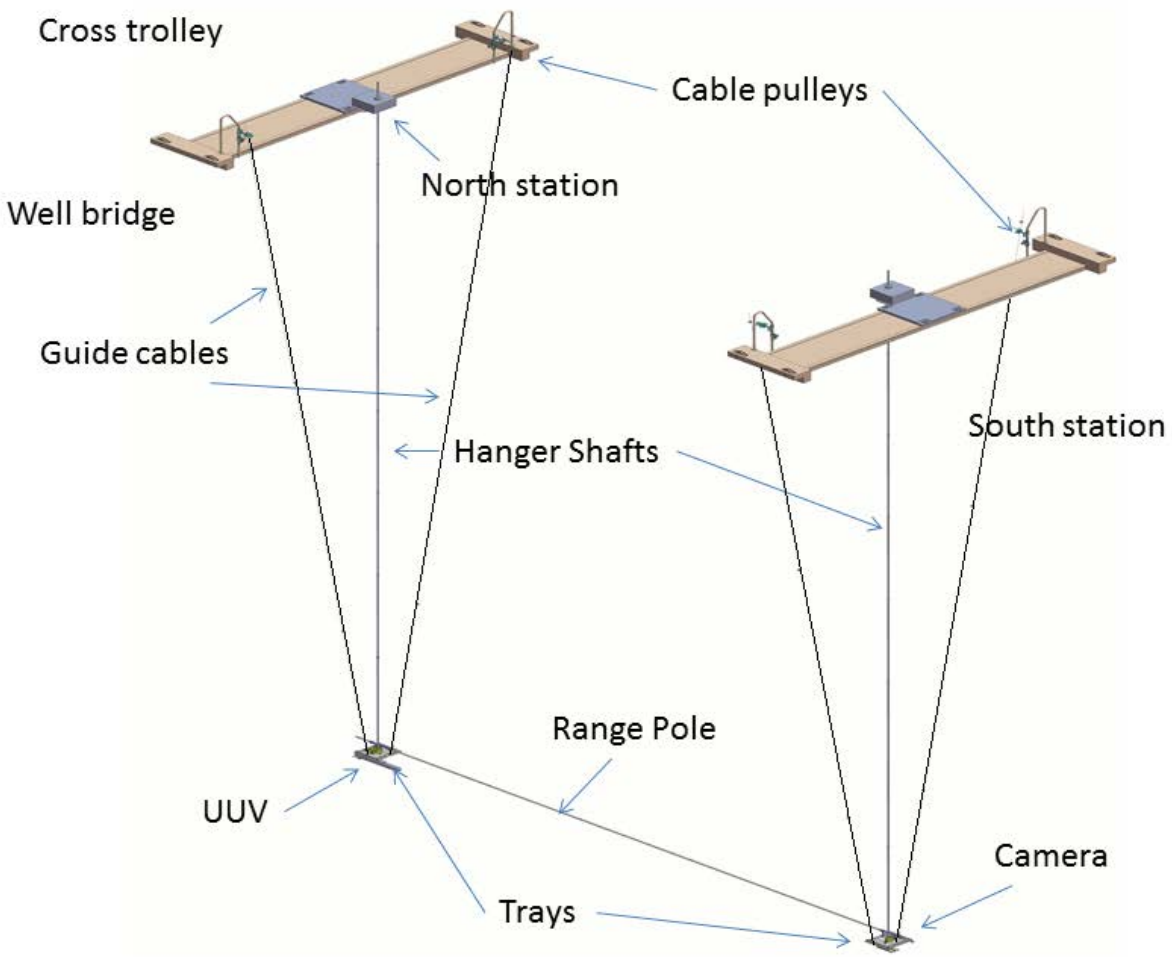
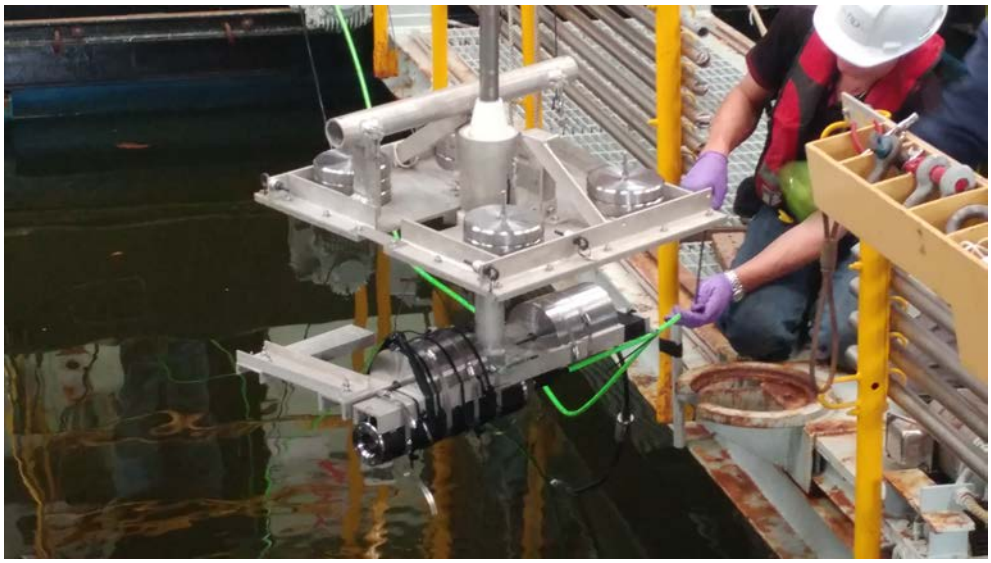


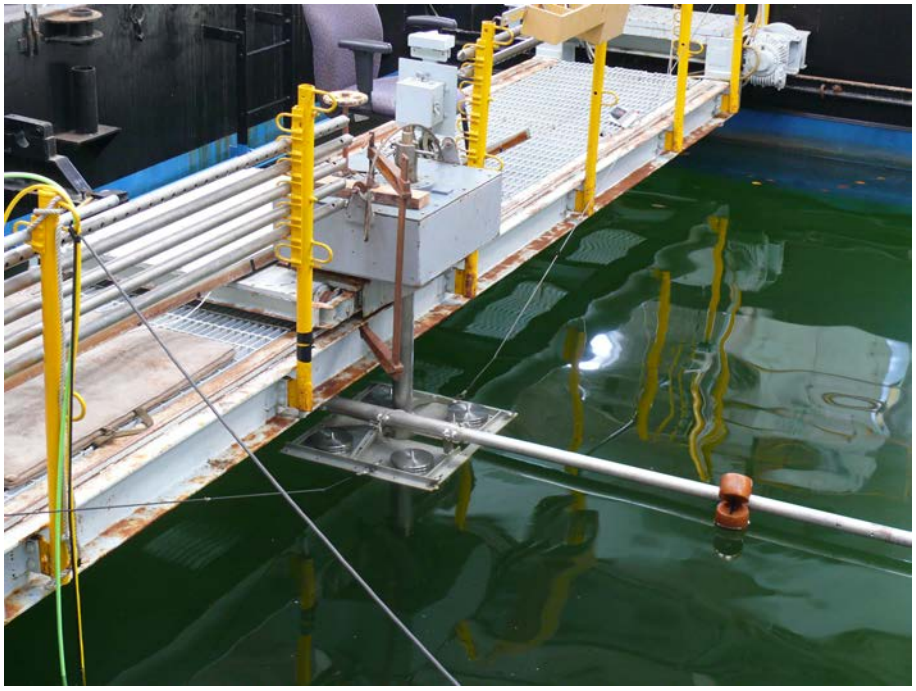
Figure 24: *An overview of the Pose Validation Rig.*



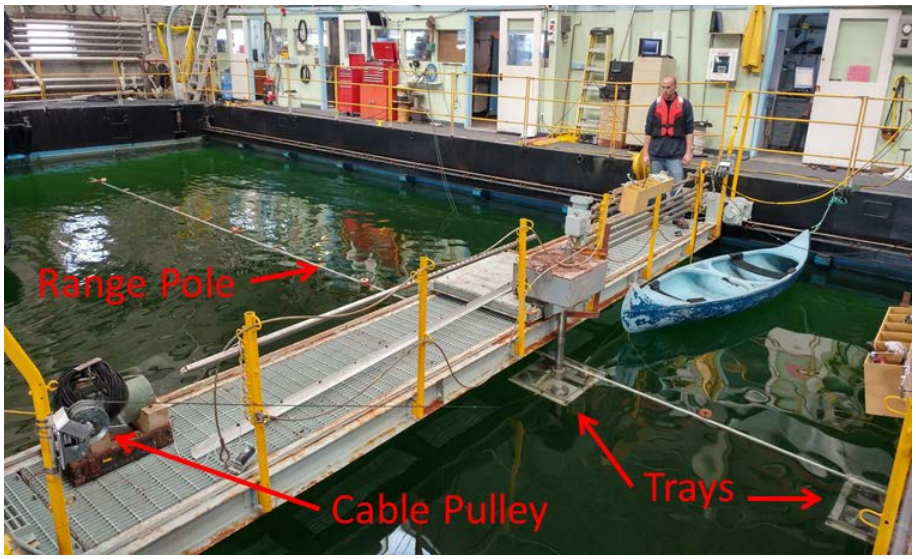
(a) *Bow station with tray and camera mount.*

(b) *Stern station with tray and surrogate UUV.*

Figure 25: *Component views of the Pose Validation Rig.*



(a) *Range pole installed on bow station.*



(b) *Range pole installed on bow station.*

Figure 26: *The installed Pose Validation Rig.*

During operation, the Pose Validation Rig functions as follows:

- The longitudinal range pole and trays rest at the bottom of the hanger shafts, maintaining the distance between the bow station (camera) and the stern station (UUV lights) using index holes drilled in the range pole at 1m increments.
- The guide cables prevent the hanger shafts from swaying in the current laterally, maintaining the relative angle between the camera and the UUV surrogate.
- The weights in the trays further stabilize the hanger shafts.
- The stern and bow station hangers are free to rotate within the trays, allowing the operators at the surface to change the yaw of the UUV. They can also change the yaw of the camera, which creates a horizontal (left or right) offset of the UUV in the camera image. The camera and UUV yaw angles are measured by encoders at the surface on the hanger shafts of each station.
- The camera or UUV may also be lowered below the trays, to create a vertical offset of the UUV in the image (up or down).

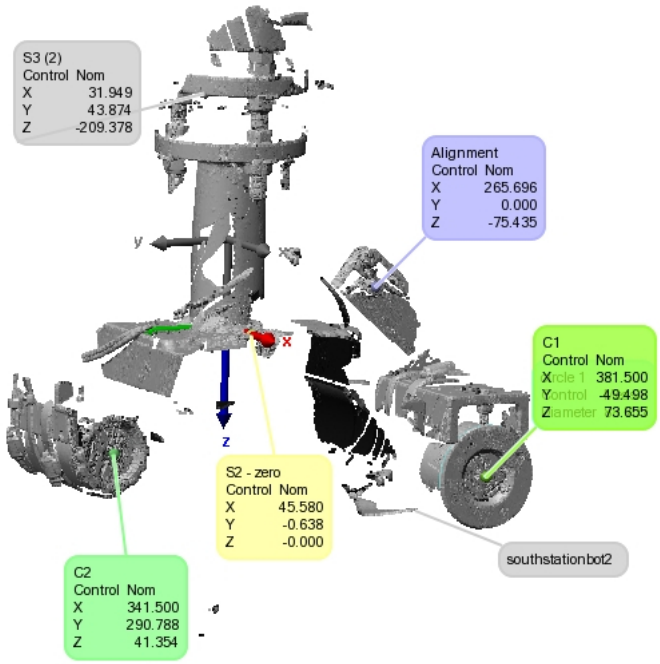
Using this system, the UUV can be positioned near or far from the camera, can be positioned anywhere in the camera field of view, and also may be positioned at any yaw, while maintaining "ground truth" measurements of their relative positions. However, as currently used, the Pose Validation Rig does not allow for UUV pitch or roll to be adjusted.

In order to change the distance (range) between the camera and UUV, the following steps are taken:

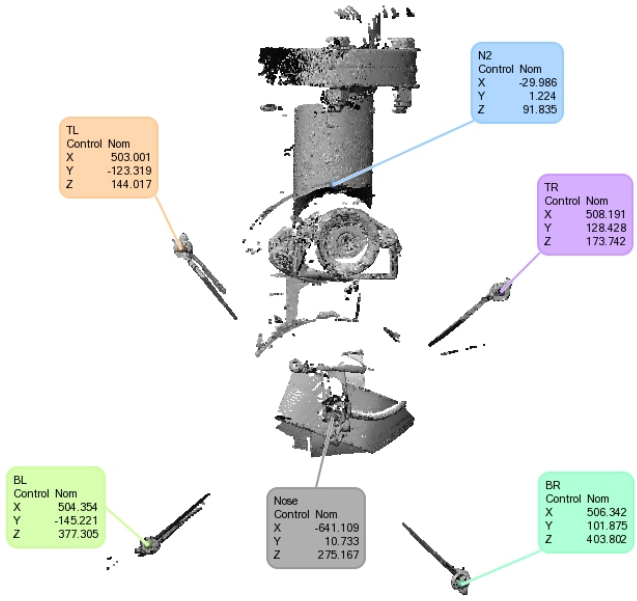
- The cable pulleys are wound by four people simultaneously, shortening the guide cables, and pulling the trays and range pole to the surface. The cables have markers on them to indicate distance, to ensure the length of each cable is the same. The trays have a slip collar around them that allows them to ride up and down the hanger shafts to the surface and back down as required. In this way the camera and UUV can remain submerged while the rest of the Pose Validation Rig is brought to the surface.
- Once at the surface, the range pole is detached from the trays, and then the well bridge can be moved to a new range. The range index pole is then reattached at the new desired range. Index holes have been drilled in the range pole at 1.0 m distance, and index pins are used to ensure the range between the trays is carefully controlled.
- The guide cables are then wound out from the pulleys, lowering the range pole and trays back to the bottom of the hanger shafts at the new range.
- Alignment cameras and LEDs at the centres of the stern and bow shafts were used to "zero" the yaw angles of both stations, to ensure that any offsets in rotations were accounted for prior to each set of data being gathered.

In order to establish the absolute positions of each of the UUV lights as well as the cameras in their mounting bracket, a laser measurement system was used to image the equipment at the bottom of the stern and bow stations prior to the trials. The laser scanner used was a Nikon MCA II Manual Coordinate Measuring Arm (Figure 27), that can be clamped to any surface and manoeuvred around to establish precise locations of points on the equipment. The offsets from the range pole positions to the camera lens and UUV lights were accounted for in establishing "ground truth" positions for comparing to the results from the computer-vision algorithm. Scans of the camera and UUV equipment are shown in Figure 28a and 28b.

Figure 27: *Scanning the bow station using the MCA II laser.*



(a) *Laser scan of the bow station camera mount.*



(b) *Laser scan of the stern station UUV mount.*

Figure 28: *Using the laser scanner to establish positions of cameras and UUV lights.*

Although the Pose Validation Rig is intended to provide "ground truth" measurements. There are some potential sources of error, which may include:

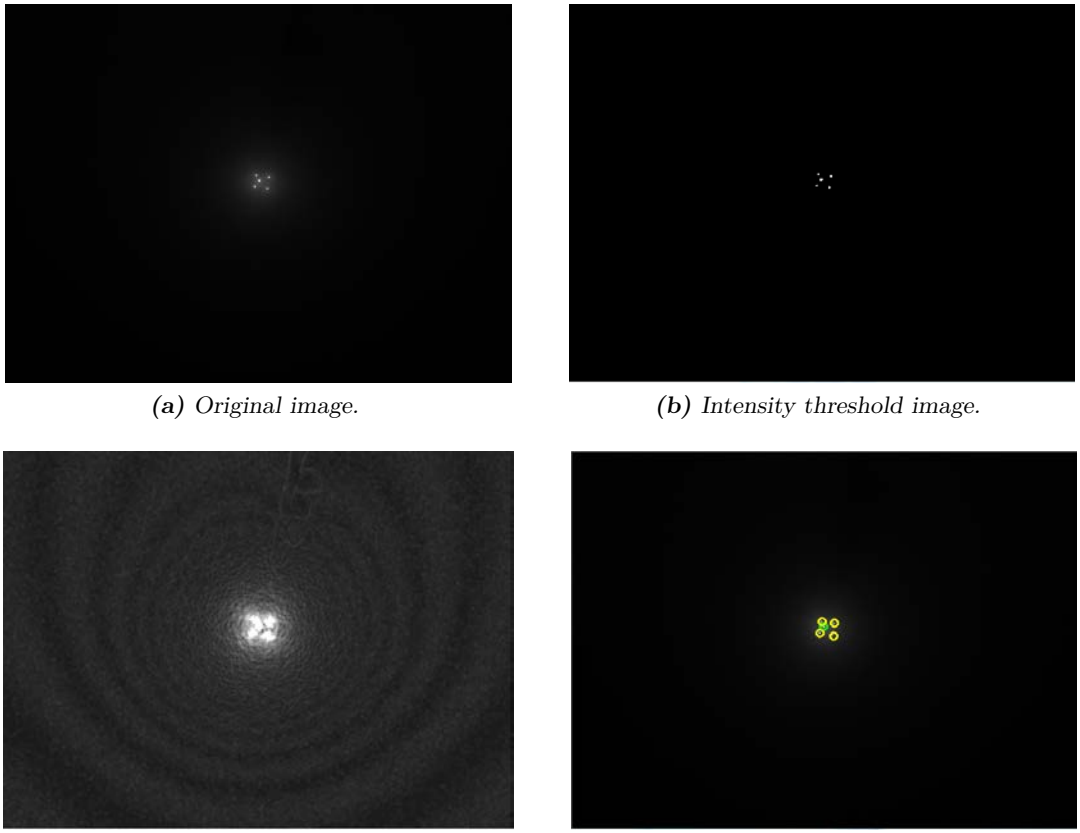
- Flexing of the range pole from movement of the hanger shafts or well bridges, shortening the distance between the camera and UUV.
- Inaccuracy in the yaw angles of both the camera and UUV. This includes error in the encoders, human error in reading the dials, and flex in the hanger shafts from waves or water currents causing unmeasured yaw.
- Lateral movement of the camera or UUV due to waves or water currents causing the guide cables to stretch.
- Height of the well bridges not being equal, resulting in either one of the camera or UUV being higher than the other. An attempt to measure this offset was unsuccessful using either a line level or a laser level.

As such, there may be some error in the data from the pose validation system. However, it will be treated as "accurate truth" in the results presented below, and will be accepted at face value for the purposes of characterizing the computer-vision system.

5 Results

5.1 Computer-Vision

In general, the computer-vision methods of identifying LED locations in the images worked well, and would not need to be adapted for a functional docking system. At 20m depth, we were able to detect the LEDs over the full range of our test facility (13m to 2m, Figure 29). If tuned correctly, the gradient method could reliably detect LEDs before a human observer looking at the video feed (Figure 30).



(a) *Original image.*

(b) *Intensity threshold image.*

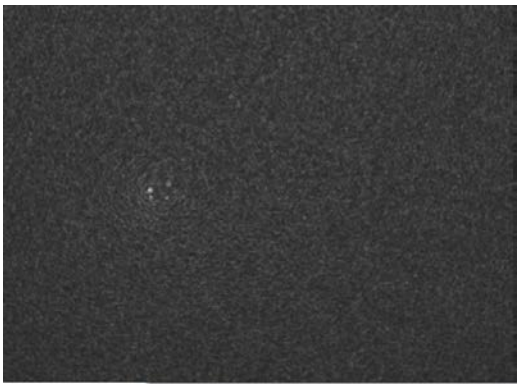
(c) *Gradient image.*

(d) *Result (gradient in yellow, threshold in green).*

Figure 29: *LED detection at farthest range tested (20m depth, 13m range, 0 degree yaw).*

The detection range would be reduced by the UUV being orientated at angles away from the camera, although this was mostly mitigated by having LED sources with a wide dispersion angle. In practice, the effective range was not significantly reduced until relative UUV orientations of 75 degrees or more (Figure 31).

(a) *Original image.*



(b) *Gradient image.*

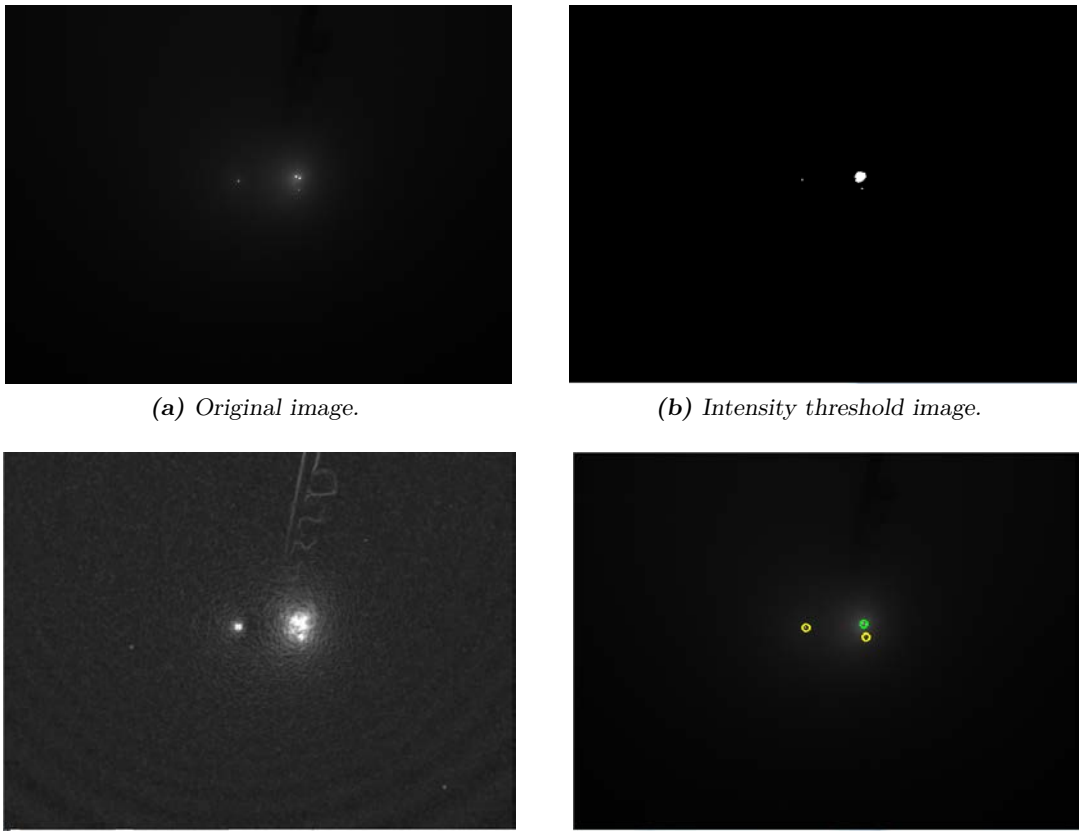
(c) *Result.*

Figure 30: *Gradient detection of an extremely dim 4 LED cluster.*

Of course, in shallow and turbid water, the detection range is greatly reduced by both ambient light and particles in the water. In the same experiments, reducing the depth from 20m to 5m reduced the reliable detection range from at least 13m to 8m (Figure 32). In turbid water, the exponential increase in absorption and scattering with range is such that increasing LED power provides only a modest increase in detection range. This indicates that automated docking will work better at depth away from sources of ambient light.

The effect of ambient lighting is particularly noticeable in Figure 33, with the UUV above the camera. In this image, the large blob at upper left is the sun shining into the water, while the UUV is in the upper right. In this case, the gradient detector still found the LED lights, but if the UUV had been directly in line with the sun, it would have been totally invisible.

As mentioned earlier, there are some problems with visual detection of LEDs. At higher UUV orientations, when LEDs begin to overlap, their centres may be found inaccurately. In highly turbid water, the LEDs may be difficult to separate, and pose information may not be obtainable.



(a) *Original image.*

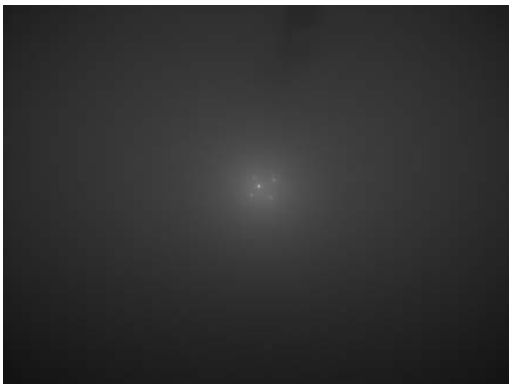
(b) *Intensity threshold image.*

(c) *Gradient image.*

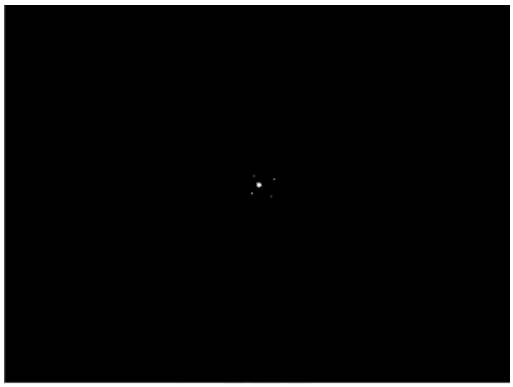
(d) *Result (gradient in yellow, threshold in green).*

Figure 31: *LED detection at farthest range tested (20m depth, 13m range, 75 degree yaw).*

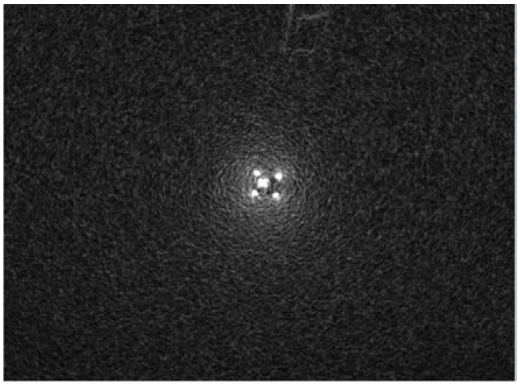
To illustrate our results, a sequence of images has been taken from a simulated docking procedure (i.e. the UUV was controlled by human operators). These images are shown in Figure 34 and illustrate the appearance of the UUV with exposure control, and the detection by the computer-vision algorithms. Note that the LEDs appear at an acceptable size for detection despite the reduction in exposure time by two orders of magnitude.



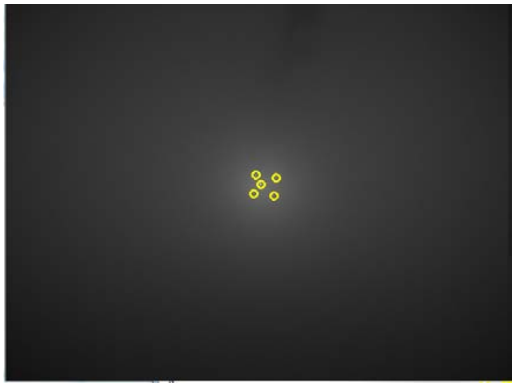
(a) *Original image.*



(b) *Intensity threshold image.*

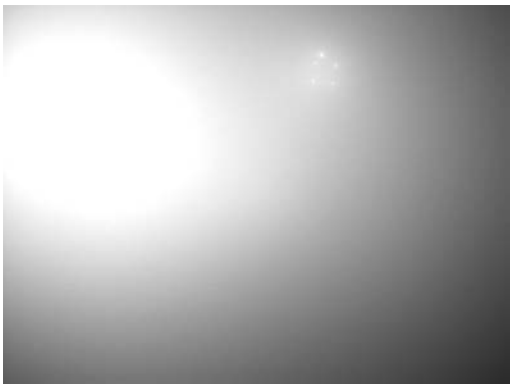


(c) *Gradient image.*

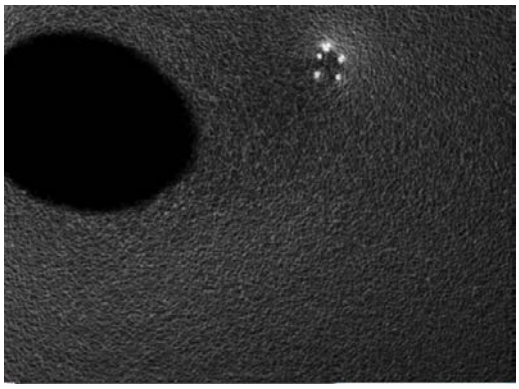


(d) *Result (gradient in yellow).*

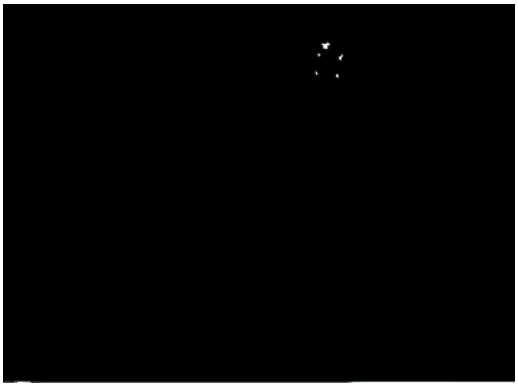
Figure 32: *Near maximum range of detection at shallower depth (8m range, 5m depth).*



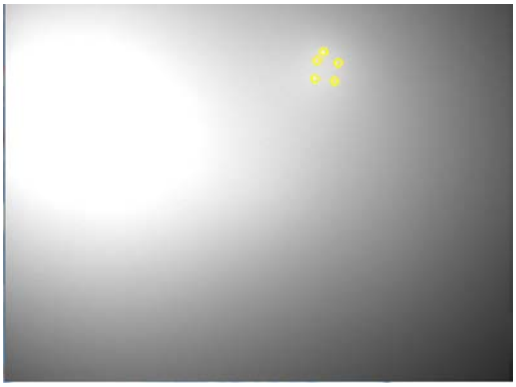
(a) *Original image.*



(b) *Gradient image.*

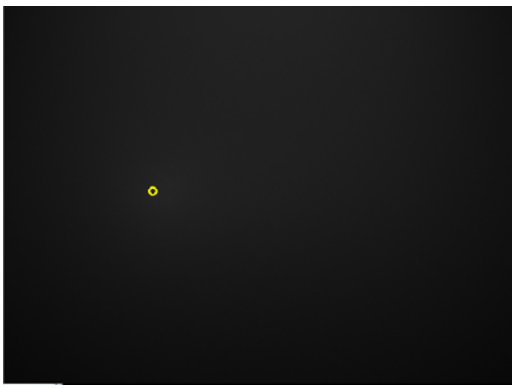


(c) *Thresholded gradient image.*



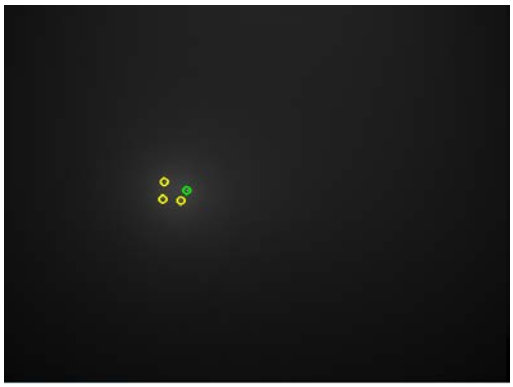
(d) *Result.*

Figure 33: *Gradient detection with turbid water and bright ambient light.*

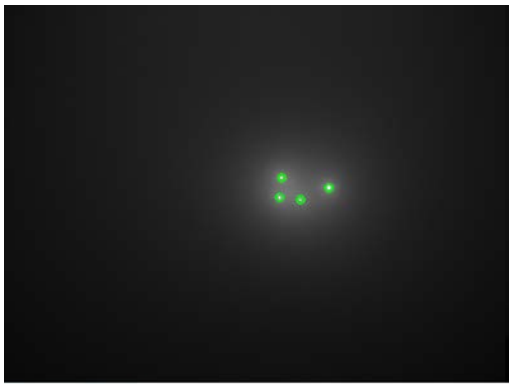


(a) *Start of acquisition.*

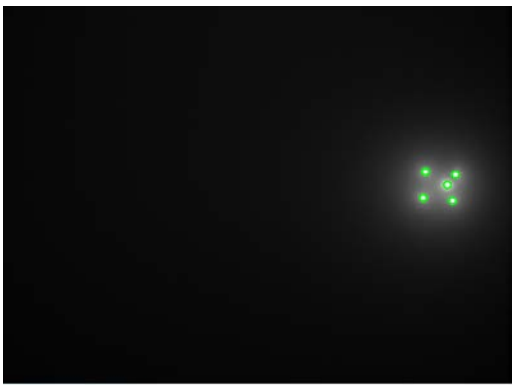
(b) *Gradient detections.*



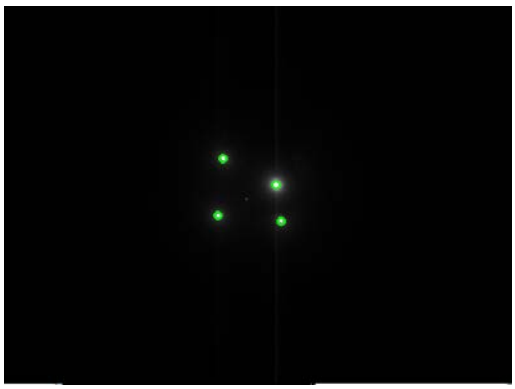
(c) *Mixed intensity threshold and gradient detections.*



(d) *Intensity threshold detections.*



(e) *Intensity threshold detections.*



(f) *Top-right LED obscured by UUV body.*

Figure 34: *Sample images from a UUV docking sequence with active exposure control from approximately 11m to 2m range. Gradient detections in yellow with intensity threshold detections in green.*

5.2 Position and Pose Estimation

5.2.1 Ground Truth Tests

Using the Pose Validation Rig discussed earlier, a set of sample images was taken of the UUV at pre-selected positions and poses. There were five different ranges tested between 2 and 13 metres. At each range, UUV yaw was varied between 0 and 85 degrees, with up to 10 different horizontal positions in the image (achieved through camera yaw), and two different vertical positions, resulting in 909 images of distinct poses. Of these, a representative set 782 predictions were selected in which duplication was minimized.

Our goal was to build the ground-truth pose validation rig system to an accuracy of ±3cm, but there is no independent means of validation. As shown in the results below, the computer-vision system and the pose validation rig are in agreement to within several centimetres of range and a few degrees of angle. This seems acceptable given the ranges involved and unknown water currents or other sources of error.

For this set of data, we found a suitable camera calibration factor (or focal length C) for predicting range by optimizing our results for range over our test data using the ground truth information. This value for C could have also been found through a camera calibration procedure or perhaps through manufacturer data.

The 6-degree-of-freedom results for some of these sets of poses is shown in Figures 35 to 40. The left side of the graphs shows the position of the UUV in spherical coordinates (range, bearing and elevation) from the camera to the centrepoint of the UUV. The right side shows the pose of the UUV, as roll, pitch and yaw. In each graph, the x-axis indicates the image number in the particular set taken at that range and depth. The blue line represents the theoretical pose based on our mechanical ground truth system, while red indicates the result as reported by the computer-vision system. Those samples where only three LEDs were visible in the image are marked with an "x". The full set of graphs for the tests with the above mentioned 782 images is included in Annex D.

For the test shown in Figure 35, the UUV was left static, and a number of images taken over several minutes to estimate the noise in the system. As can be seen, the bearing and elevation estimates are the most stable and accurate because they are based only on the overall position of the LEDs in the image, not the position relative to each other. In addition, there are 17 pixels of resolution per degree of bearing/elevation, so it is relatively easy to achieve accuracy. By contrast, the range, roll, pitch, yaw estimates are based on the relative positions of the LEDs to each other, and are therefore a much smaller number of pixels in measurement, and are therefore much more noisy. In addition, there is an offset in the pitch measurement that was present in all of the results, for which we do not have an adequate explanation, but it is likely related to the physical setup of the ground truth measurement system.

Figure 36 shows the results for a test where the UUV was left in the same position, except now it was rotated in yaw. Similar results are obtained, except when the UUV has yawed

away from the camera by about 75 degrees or more and only 3 LEDs were visible. At this extreme angle, due to lack of a visible LED on the back side of the hull, roll calculation was the most affected.

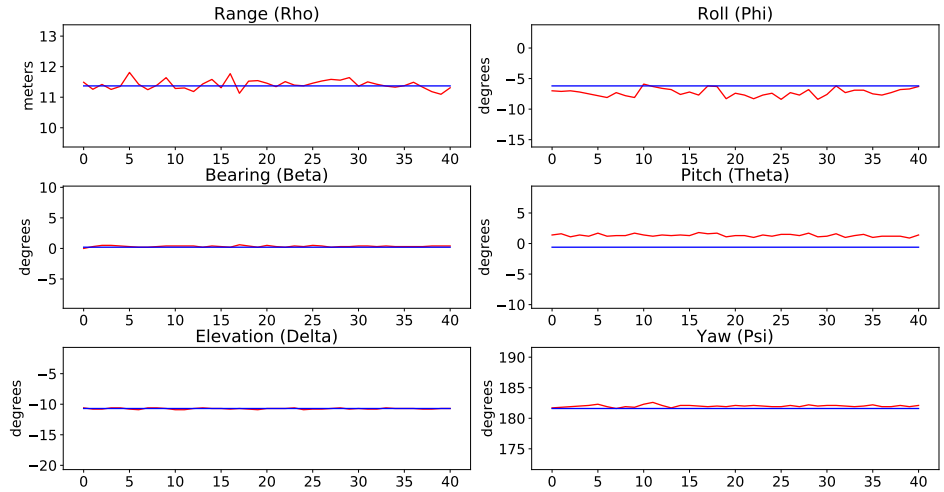


Figure 35: *Pose results from a set of images captured at 11m range and 20m depth, without the UUV pose or position changing.*

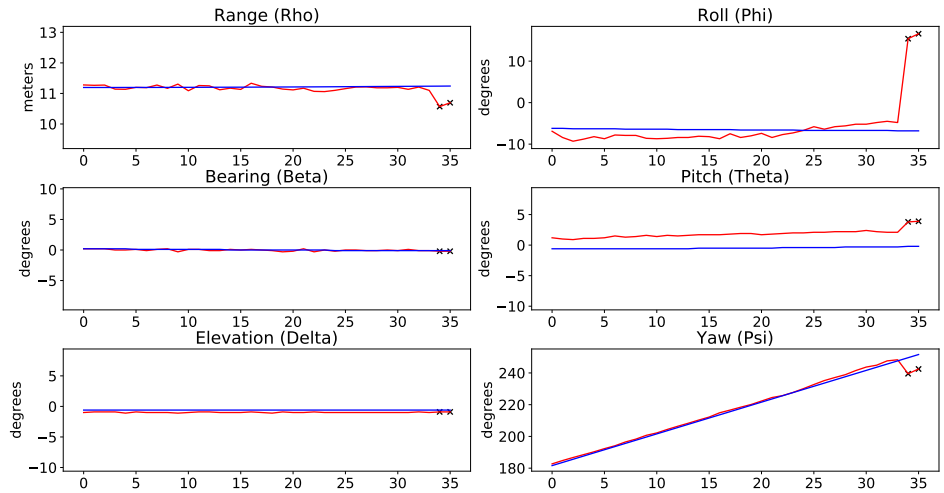


Figure 36: *Pose results from a set of images captured at 11m range and 20m depth, with UUV static except for a rotation in yaw.*

Figures 37 to 40 show results where the UUV was moved around in the camera field of view. These tests were conducted with the range specified held static by the test rig, and consisted of moving the UUV horizontally across the images, while also changing its yaw heading.

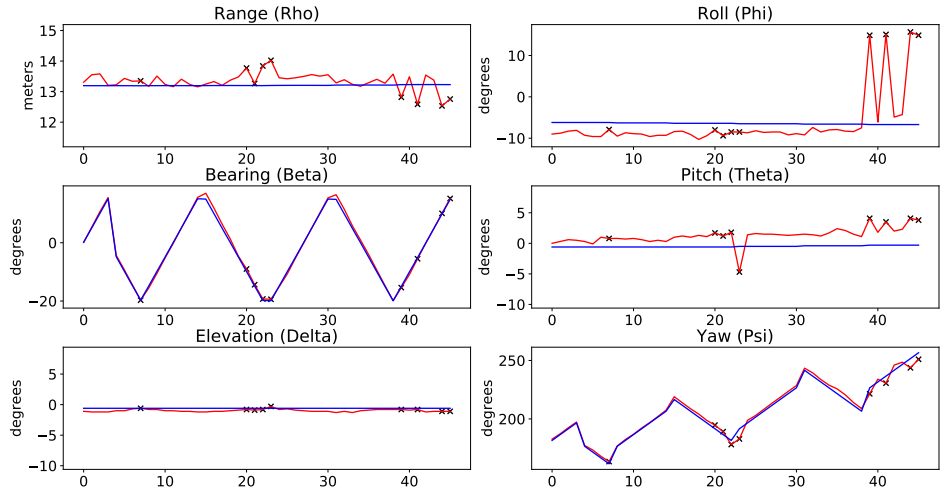


Figure 37: *Pose results from a set of images captured at 13m range and 20m depth.*

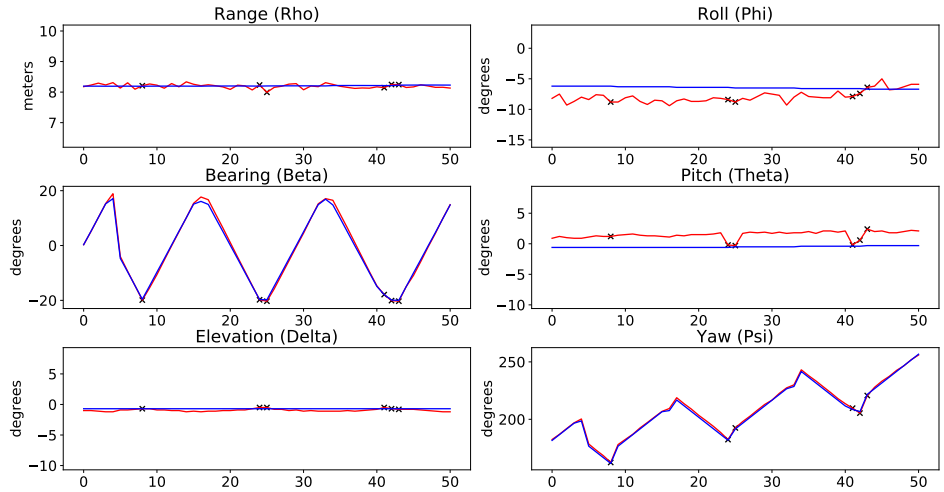


Figure 38: *Position/pose results from a set of images captured at 8m range and 20m depth.*

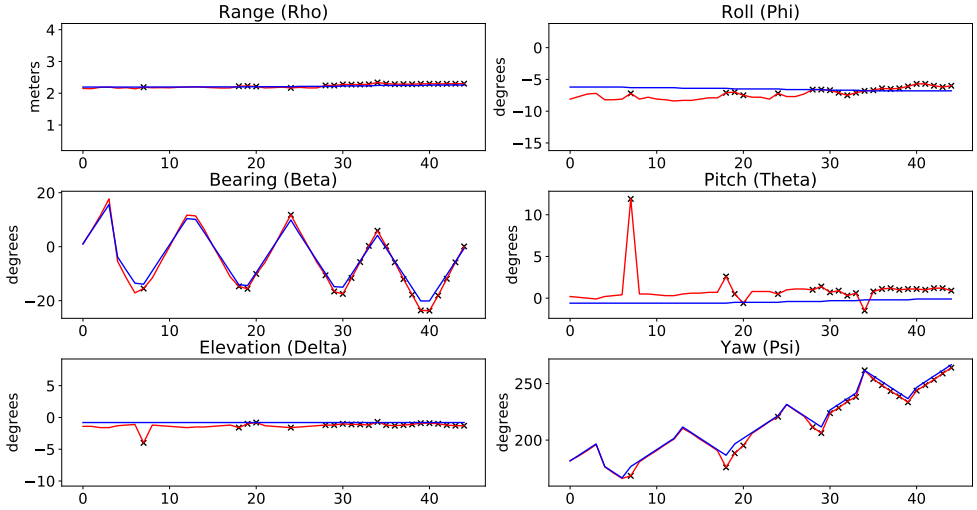


Figure 39: *Position/pose results from a set of images captured at 2m range and 20m depth).*

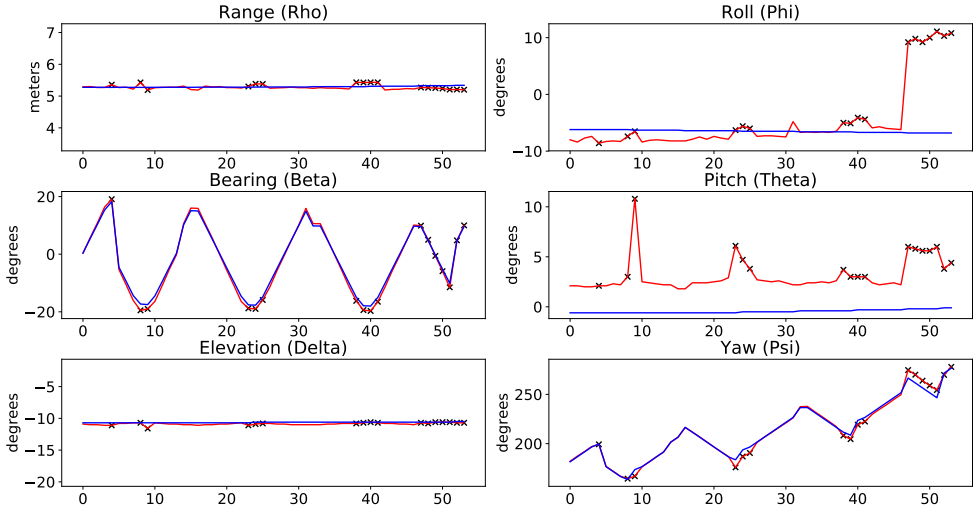


Figure 40: *Pose results from a set of images captured at 5m range and 20m depth. This test also had an offset such that the UUV was situated in a position above the horizontal plane of the camera by approx. 1m.*

A summary of the range accuracy over the set of 782 poses is shown in Figure 41. As expected, as range increases, the accuracy in predicted position decreases. At longer range, there are fewer pixels between each of the LEDs to use as a measurement baseline for range estimation.

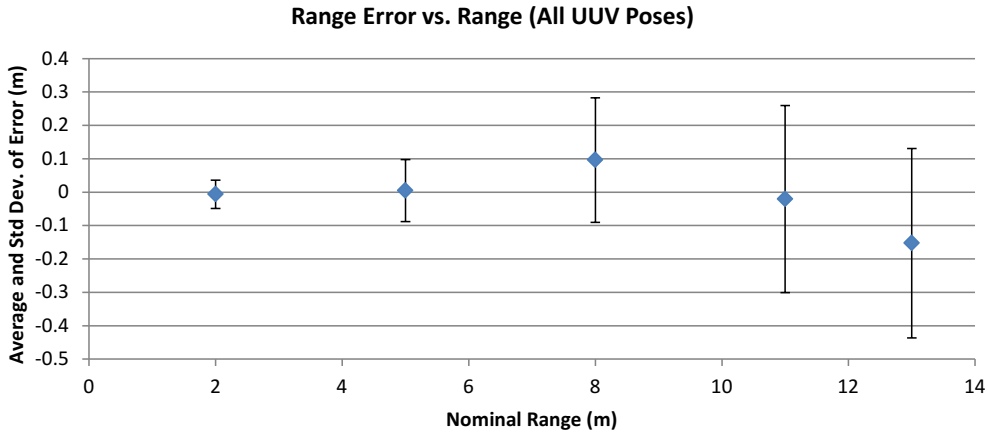


Figure 41: *Average and standard deviation of error in range estimation.*

The resulting average distance between the estimated position of the UUV and the "ground truth" position of the UUV is shown in Figure 42. This error takes into account the error in range, bearing and elevation which combine to provide the spherical coordinates of the UUV position. The position error is the Euclidian 3D distance between the predicted and actual pose of the UUV:

$$\text{Position error} = \sqrt{{x_{error}}^2 + {y_{error}}^2 + {z_{error}}^2}$$

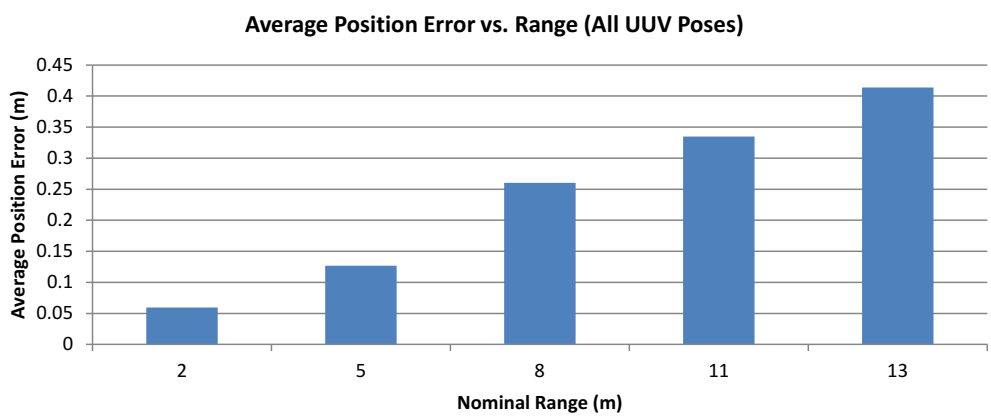


Figure 42: *Average position error over all poses.*

Higher relative UUV yaw also decreased accuracy in range, although less dramatically (Figure 43). With UUV yaw, the distance between nose and tail LEDs increases, which should improve range accuracy. However, with higher yaw, the LED positions from the computer-vision software was less accurate, fewer LEDs are visible, and the spacing between tail LEDs is smaller, resulting in poorer estimation of UUV pose angles (Figure 44). As such, roll is poorly predicted when we can't see the LEDs on the back side of the UUV hull. Because all of the angles are interrelated in the nonlinear S function, if roll is miscalculated, all the other angles will have a poorer result as well.

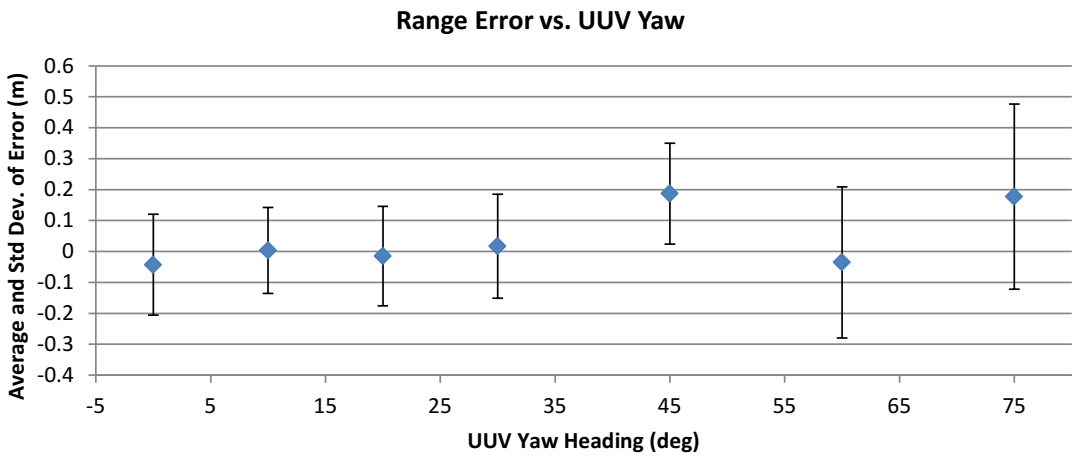


Figure 43: *Average and standard deviation of error in range estimation with increasing yaw over all poses.*

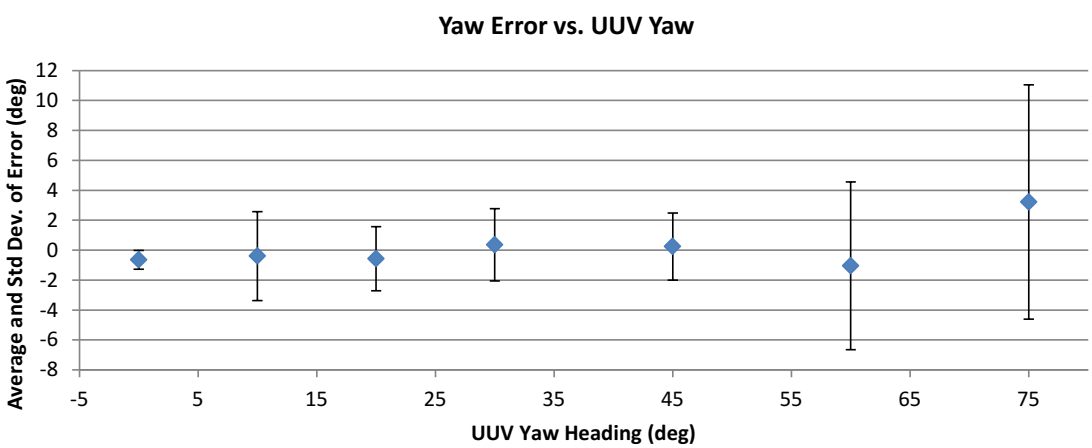


Figure 44: *Average and standard deviation of error in yaw estimation with increasing yaw over all poses.*

Overall statistical results from the set of 782 poses are shown in Table 1. Accuracy improves with more LEDs, although this is somewhat overstated because the poses where more LEDs

are visible are less eccentric (i.e. the UUV is located more in the centre of the image without significant yaw when 5 LEDs are visible).

Table 1: *Standard deviation of 6-DOF error over all poses.*

	3 or more LEDs	4 or more LEDs	5 LEDs
Range/ρ (% of measurement)	2.12	1.30	0.58
Bearing/β (deg)	1.28	1.26	0.56
Elevation/δ (deg)	0.35	0.25	0.09
Roll/ϕ (deg)	4.60	1.19	0.49
Pitch/θ (deg)	1.37	1.03	0.50
Yaw/ψ (deg)	3.77	1.29	0.38

5.2.2 Simulated Docking

In order to try to understand more about the performance of the pose estimation system, a "simulated rendezvous" was performed by continuously moving the UUV surrogate from the farthest range of the facility to the nearest. While doing this, the surrogate was also yawed and translated side to side to simulate UUV motion. The simulated rendezvous was subject to the limits of the barge facility, with UUV pitch and roll remaining fixed during the experiment.

For these tests, the pose validation rig was not able to be used. However, the results demonstrate the amount of noise in the pose measurements while moving, as well as the conditions under which problems occur with the computer-vision and pose estimation.

Some typical results are shown in Figure 45, with some sample images from this test shown in Figure 34. The system begins to provide reliable pose estimates as soon as multiple LEDs are visible, at a range of about 10.7m.

This test was conducted at a depth of only 3m, in somewhat turbid water, with the LEDs operated at the brighter end of their range (750mA operating current). If conducted at depth, the range achievable is expected to increase. However, even under these conditions, reliable pose estimates were achievable almost immediately when multiple LEDs were detected. As expected, at longer ranges, and when only 3 LEDs were detected, the accuracy suffered.

There was also some outlier data, that typically occurs when the 5th LED that can be obscured by the UUV body either appears or disappears. This was because when the LED first started to appear, the centre of the light was not visible, but the projection of the light in the water was. At this point, the computer-vision system inaccurately founds the LED centre, causing inaccurate pose measurements. Pose filtering using a Kalman filter or other method would easily eliminate these outliers.

The detailed 6 degree of freedom plots are shown in Figure 46. In addition to noise, there are some offsets in the measurements. Most noticeable was the UUV roll, which was due to the mounting angle of the UUV (measured at approximately 5.4 degrees by the laser measurement system). There also is a slight offset in elevation, meaning that the camera and UUV were not quite level in the water, which we had difficulty removing in the test setup.

Another simulated docking test was also conducted with LED current at only 5mA, more than two orders of magnitude less, which would be more practical for a battery powered UUV. Results are shown in Figure 47 and 48. In this case, the increased exposure of the camera was able to compensate for the reduction in LED brightness. However, at these shallow depths UUV detection range was decreased from almost 11m to 7.3m.

In this test, the data outliers caused by the appearance and disappearance of LEDs can also be seen. Even at close ranges, the mis-location in the image of an LED source is by far the largest source of error.

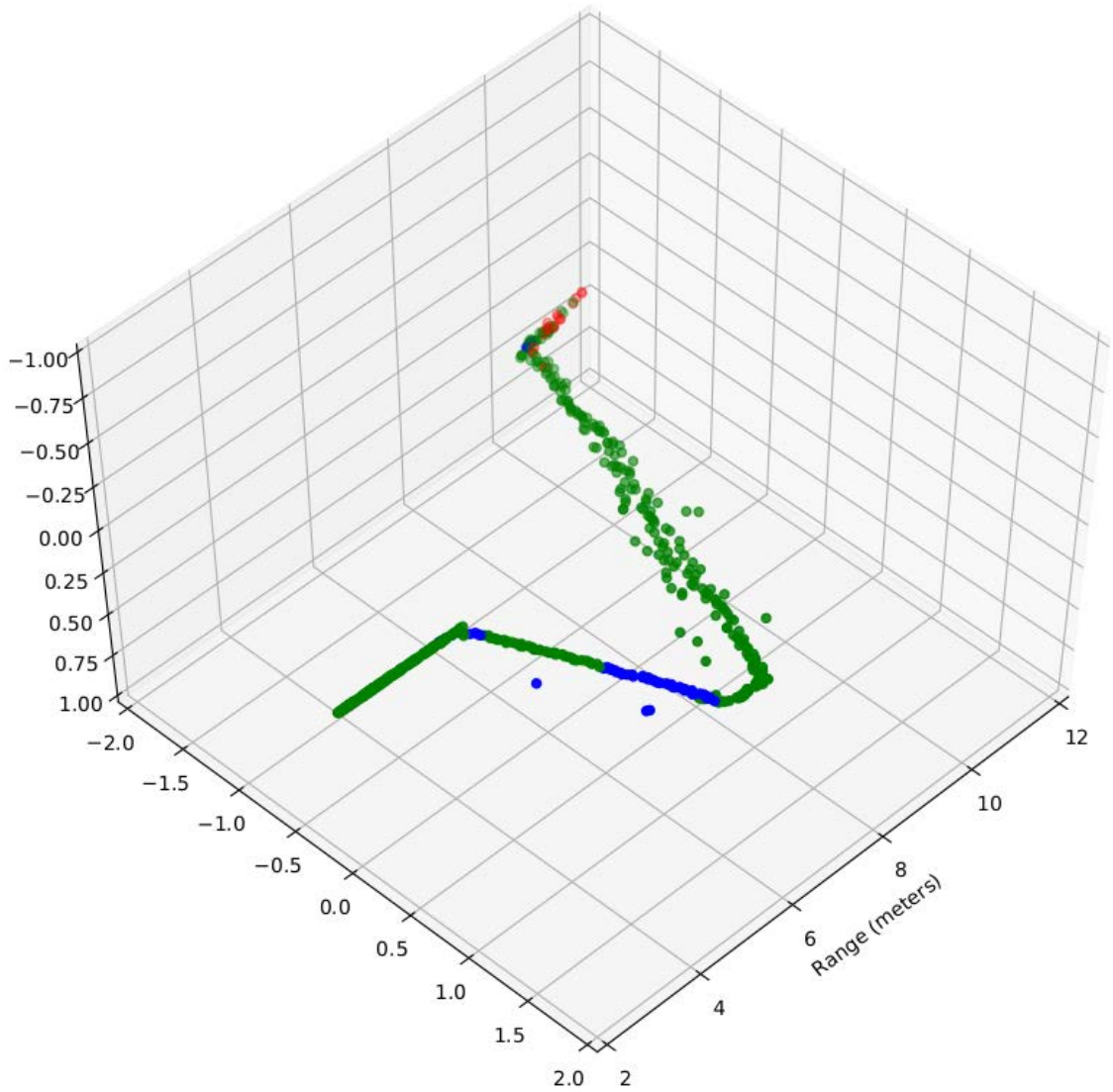


Figure 45: *3D plot of a simulated docking manoeuver. Images with 3 LEDs detected are shown in red, 4 in green, and 5 in blue.*

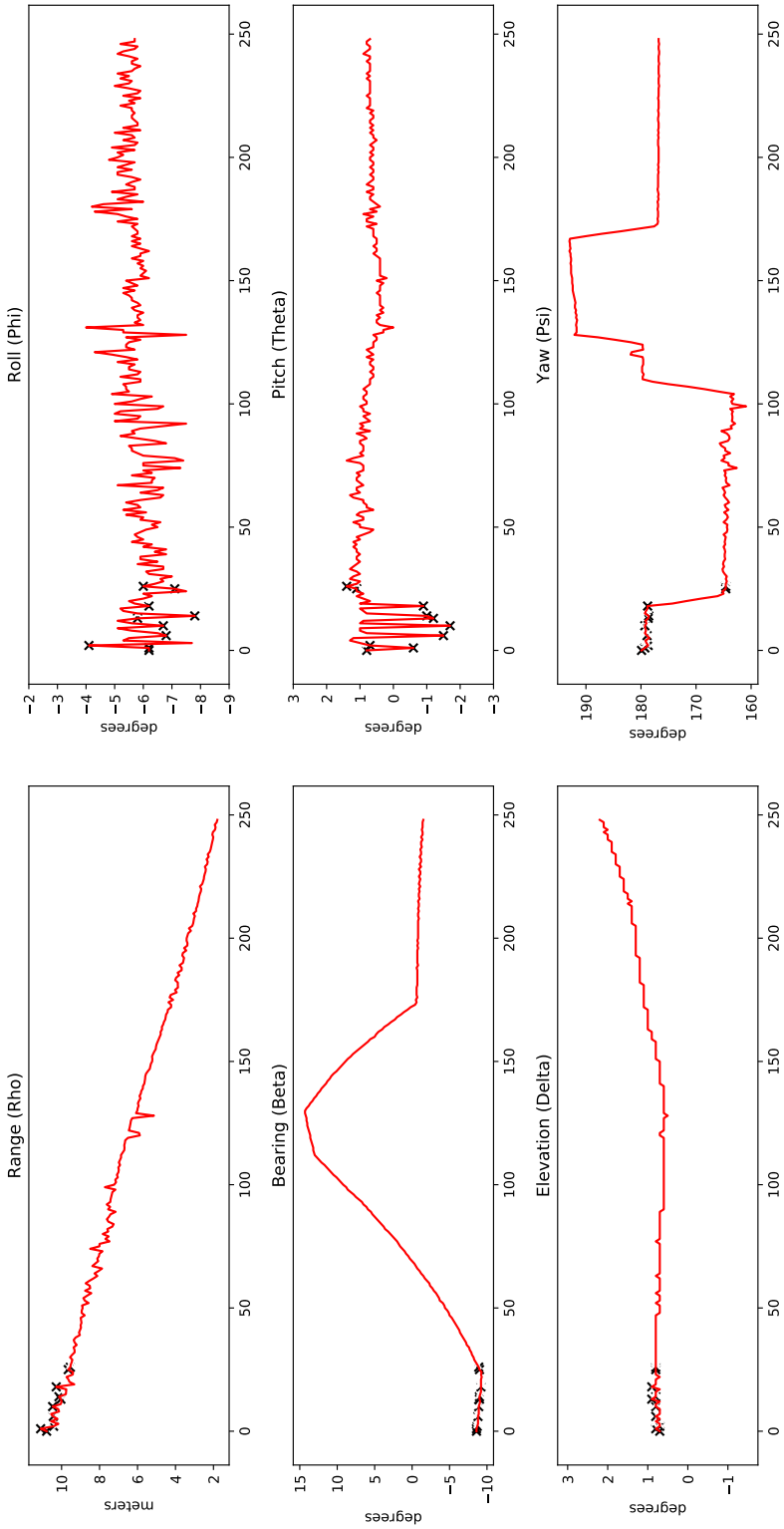


Figure 46: *6 DOF plots of UUV pose during simulated docking. Images with only 3 LEDs visible are indicated with an "X".*

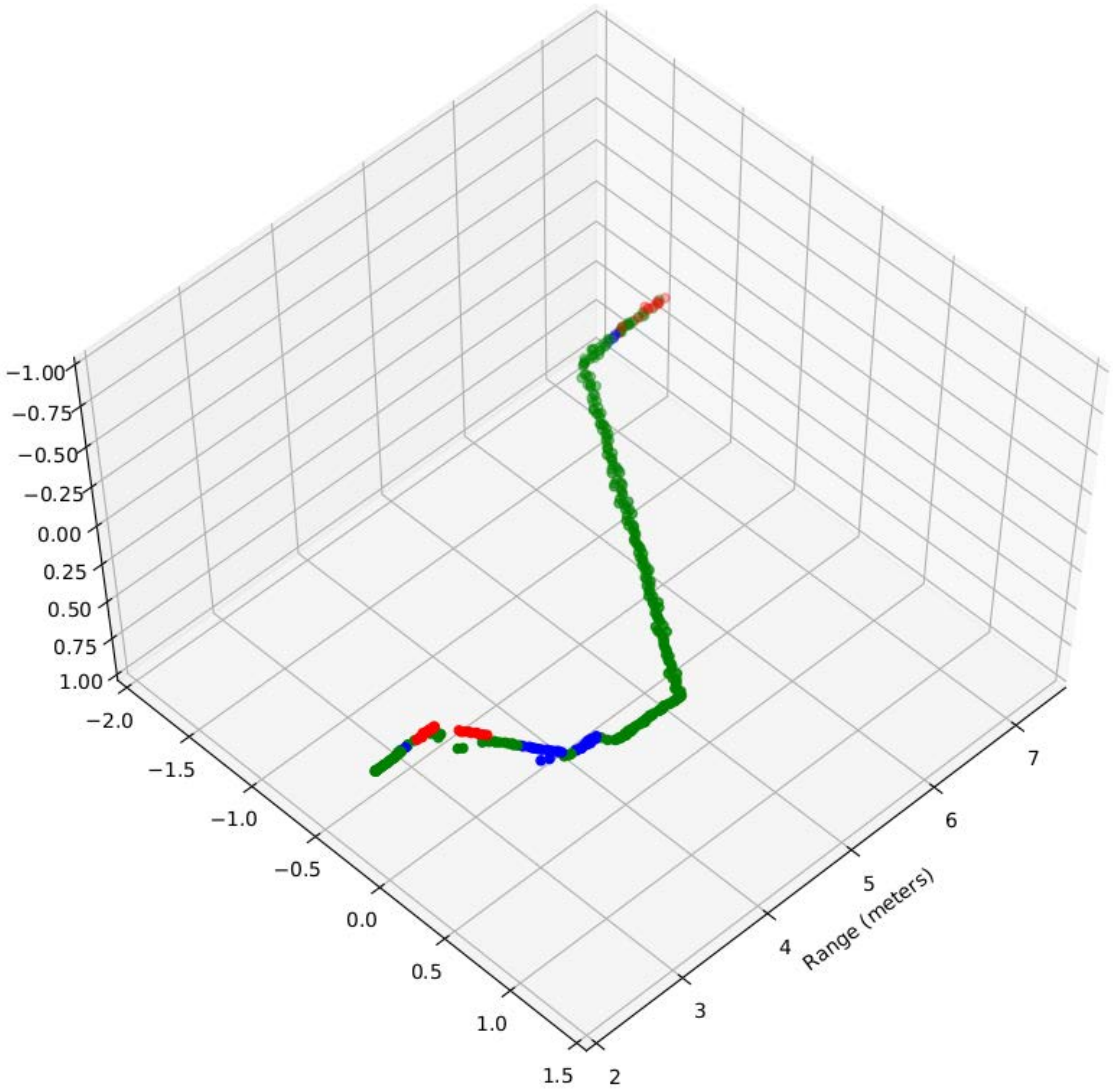


Figure 47: *3D plot of a simulated docking manoeuver with low power LEDs (5mA). Images with 3 LEDs detected are shown in red, 4 in green, and 5 in blue.*

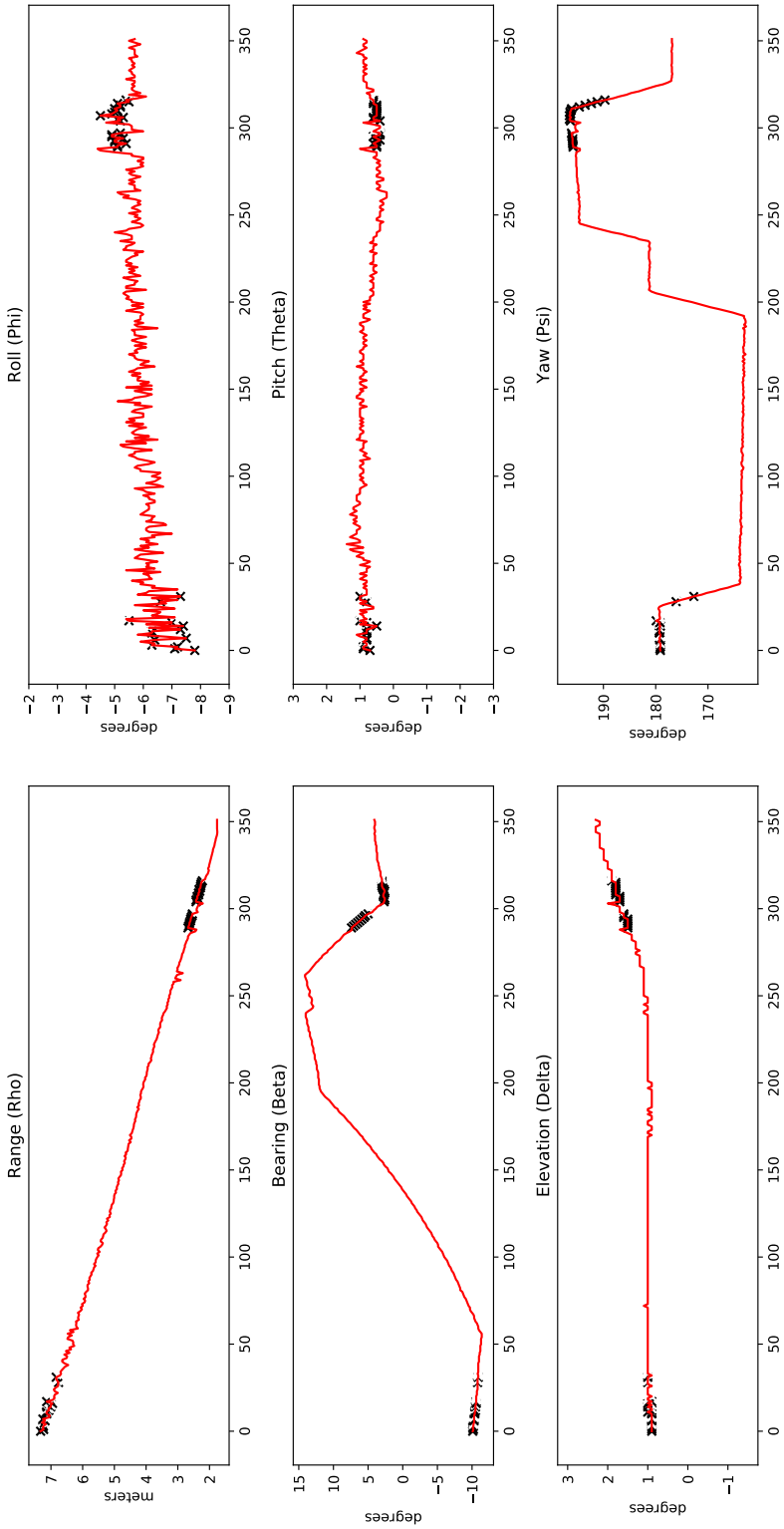


Figure 48: *6 DOF plots of UUV pose during simulated docking with low power LEDs. Images with only 3 LEDs visible are indicated with an "X".*

6 Discussion

Despite the encouraging results obtained, there remain improvements to be made. With respect to the pose estimation of the LED, no Kalman filter or other multi-hypothesis method were used. This would help to remove noise from the estimation system, and reduce errors when fewer LED lights are visible in the image.

We have learned that, in general, the effect of a reduction in ambient light and increase in water clarity is to improve the range and accuracy of the system. As such, the docking system would be best used at depth, rather than at the surface.

Despite the demonstration of effective pose estimation using multiple light sources, it is also possible that for most docking scenarios, the system could be effectively controlled by simple "homing" guidance. This is particularly true since the bearing and elevation to the target are the most accurate measurements. Such a system could work with a camera placed near the dock end effector, and by keeping a single nose UUV light in the centre of the field of view. In such a scenario, the image processing and exposure control demonstrated here would still be a critical part of the system. However, even with pose estimations, we have shown that pose and range estimates improve as the UUV gets closer to the camera, providing another reason for the camera to be as close as possible to the dock end effector.

It would also be possible to augment this single camera solution. Stereo pairs of cameras would provide improved range measurement capabilities, although probably only at shorter ranges [18]. It might also be possible to extend the stereo baseline (and range) by placing a second camera farther away on the dock. In the extreme, an extra camera system could be used on another location on the sub, providing triangulation of UUV position, and greatly improved accuracy. It is imagined that extra outward facing LEDs on the side of the UUV would aid in accuracy.

The addition of a zoom camera has the potential to improve the range of this system as well, if the focal lengths were carefully controlled. Such a system could initiate with a wide angle for detection, then zoom in on the target at long ranges to attempt to identify individual LEDs and UUV pose.

It remains for us to test this system against a live UUV manoeuvering realistically, as opposed to one mounted on a slowly moving test rig. It should also be tested as part of an active dock to understand how the movement and dynamics of a real scenario would affect performance.

6.1 Camera Calibration

When conducting experiments and gathering data, there was some concern about the need to calibrate the cameras and remove distortion from the resulting images. Calibration is a common practice for many computer-vision applications interested in 3D geometry. In the

UUV docking application, calibration can serve two purposes: the first is to find the camera parameters (particularly focal length calibration factor C) for accurate position calculations of the target. Secondly, the resulting data (particularly distortion parameters) can also be used to un-distort images to further improve accuracy.

As such, several dozen images were taken in a freshwater tank at DRDC – Atlantic Research Centre with a 20x20 inch checkerboard, with 10x10 squares at a wide variety of poses at distances of 1 to 3 metres (Figure 49). The images were then processed with the OpenCV camera calibration utilities. The results of this calibration were not successful, with the software reporting a high re-projection error. In addition, the calibration factor (focal length) found by the system would vary dramatically depending on which set of calibration images were used.

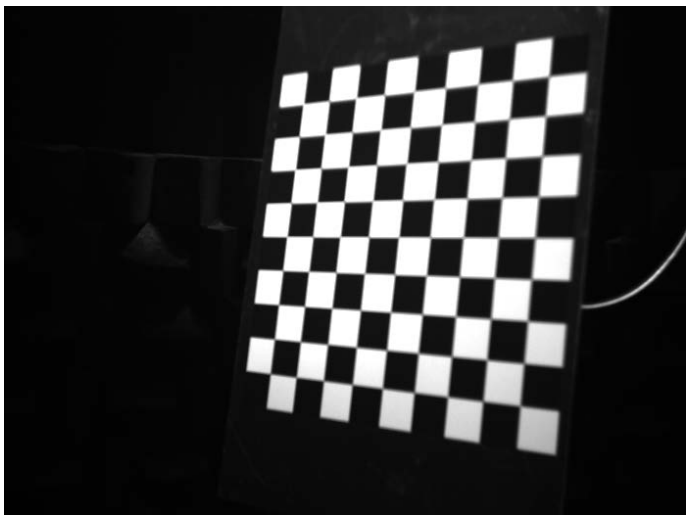


Figure 49: *Checkerboard used for camera calibration.*

The resulting failure of calibration is almost certainly due the light refracting through the glass pressure housing on the exterior of the camera, making the perspective camera model used in this calibration software invalid. However, as shown in Annex C, the error introduced is quite limited. If required, there are methods of calibrating cameras underwater [30, 31], but given the accuracy achieved in our measurements, we have deemed it beyond the scope of this project.

To our benefit, most of the images are taken at long distances, making the separation between the LEDs on the UUV a small percentage of the image. This means that any error introduced by distortion would be minimal because of the few numbers of pixels involved. However, inaccuracies in the estimation of bearing at the edges of the images can be seen in Figures 37 to 40, almost certainly due to image distortion. At shorter distances, distortion may have been a factor in introducing range measurement error. However, at short distances, any percentage error in the range measurement would be less in absolute terms, and is therefore of less concern. It also does not seem to have affected the algorithm's

ability to come to solutions for pose estimation.

This does leave the question of an appropriate focal length (calibration factor C) to use for our position estimation. In our case, we optimized the calibration factor to get the best overall match to our ground truth data. In the future a proper calibration accounting for refraction by the glass pressure housing should be used.

6.2 LED Positioning

LED positioning on the UUV is also important for pose estimation. It is beneficial to have as much separation between the LEDs for maximum accuracy of pose estimation. It is also important to position the LEDs where they can be seen from the widest variety of angles and not be obscured by the UUV body (i.e. right at the nose, and on the tail planes as far as possible away from UUV).

It is also beneficial to have as many lights on the UUV as possible, as pose detection performance improves with more data points. More LEDs also means that more will be visible at any given time despite some being obscured by the UUV body. However, it is also important to have separation between the LEDs in order to prevent the computer-vision system from identifying them as one single light at long ranges and in turbid water. All of these concerns must be factored in when choosing LED locations.

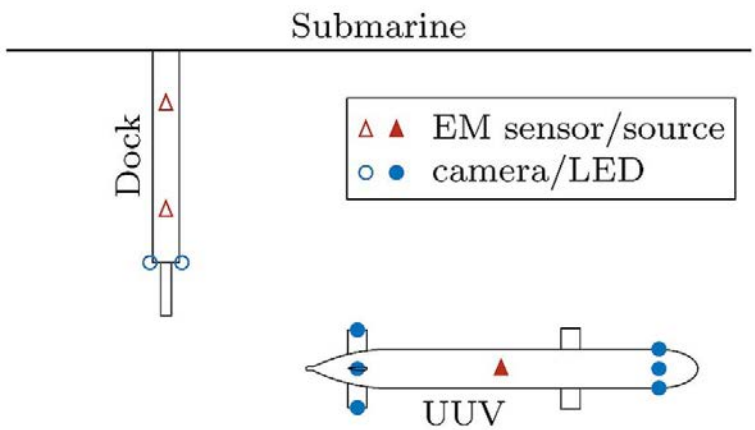


Figure 50: *Possible LED configuration for docking.*

The results shown above only used one geometry of LED lights, but it would be worthwhile to test other geometries with more lights in different positions to try and maximize the performance. Optical position and pose sensing will work best if camera and light placement are optimized for the UUV manoeuvring strategy. As shown in Figure 50, a more complex docking scenario would allow the UUV to enter the docking envelope forward of the dock and allow the dock to gradually overtake it. As docking proceeds, the UUV could end up either forward or aft of the dock at any given time, depending on whether nose or side

capture is used or if capture attempts fail. Camera placement should be inboard of the outer dock link. Two cameras are shown but there could be more, enough to cover about a 180 degree sector off the end of the first link.

Our results so far use lights facing forward for nose-on docking. However, there is no limitation on placing LEDs on the side of the UUV facing out, and on the rear of the UUV facing backwards. With low bandwidth communications, the dock mechanism could communicate with the UUV to indicate which LEDs to activate and deactivate for its current position relative to the dock.

6.3 LED Characteristics

When selecting LEDs, it is beneficial to have as wide of a viewing angle as possible with even luminosity (i.e. a diffuse light source). This provides even brightness at all UUV poses, allowing the LEDs to be visible at a wider range of UUV headings. It also means that less exposure control is required to accommodate UUV heading changes.

Additionally, our LED sources were powered at only 100mA for most of these tests (with a possible maximum current of 1 Amp). With higher current, and in clear water and at lower depths, the LEDs could probably be detected much farther. This would probably require an active control system on the UUV to adjust LEDs to be strongest at far ranges and lower at near ranges to avoid saturating the image sensor. At long ranges, the pose estimation would become problematic because of the lack of separation between the LED sources in the image, but it would however allow for early detection of the UUVs relative bearing and declination for early correction.

A more sophisticated approach to rejecting ambient light could be feasible through the use of laser diodes radiating at a particular frequency in the blue or green spectrum. This wavelength could then be selected for using bandpass spectral filters on the camera to reject ambient light and target the particular diode light. Diffusers on the laser diode could provide for the necessary viewing angles.

6.3.1 LED Identification

During the previous sections, there was no mention of the method of determining which LED is which from the images. This is an important point, as the pose solution relies completely on this information to avoid ambiguous pose solutions. It is especially critical given that there are many UUV poses which may obscure some of the LEDs from view. So far, we have made the assumption that this problem would be solvable with existing technology, of which there are a few possibilities.

The first would be designing the geometry of the LEDs on the UUV to provide clues. For example, pairs of LEDs on each of the horizontal tail plans could be used to identify them from the nose or vertical tail planes (Figure 51). In addition, it would be possible to

make assumptions about the UUVs pose, such as limiting the roll, yaw and pitch relative to the camera to aid identification. Cues from the UUV motion while being tracked over a sequence of images could also provide information for the UUV's direction of travel, in order to identify nose from tail LEDs.

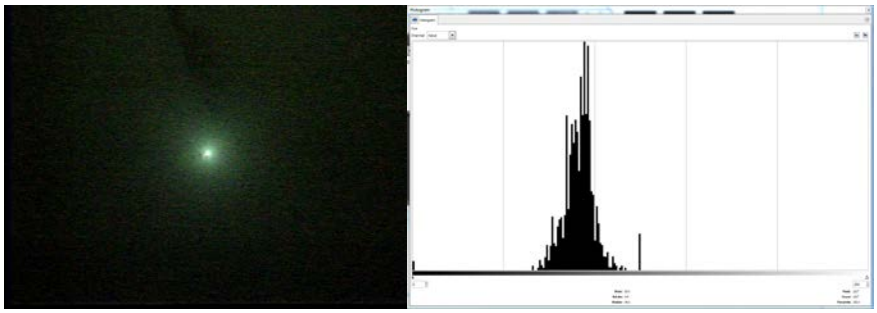
Figure 51: *Sample image with dual LEDs on each tail plane for identification.*

A second option would be to use synchronously flashing LEDs, with each timed to blink at a certain rate [32, 17]. If used with a camera system that captured images at a predictable rate, the appearance or absence of LEDs in sequential images would identify them. This would be feasible underwater because it can be assumed that the blinking rate and camera capture rate could be made much faster than the dynamics of the UUV and capture mechanism would require for stable control.

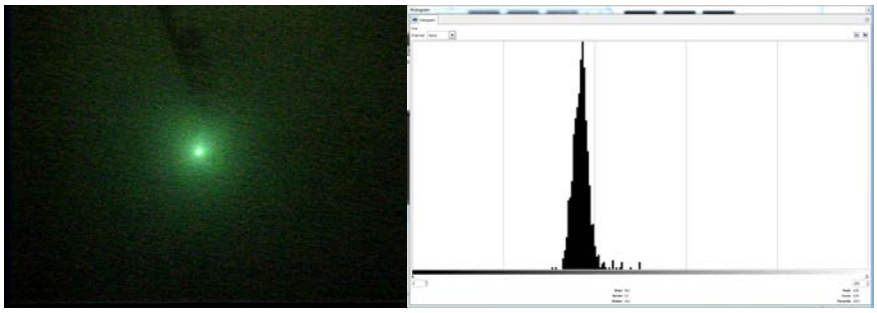
We have also briefly investigated the possibility of using coloured LEDs for identification. An unfortunate aspect of sea water is the tendency to act as a band-pass filter, turning every light source green or blue depending on the contents of marine life, silt, etc. in the water. We captured several images of white, green and blue LEDs at several distances to examine the feasibility. The test conditions on these days involved relatively turbid water with runoff from a nearby river and recent rains.

The images, and histograms of their hue data (from HSV colourspace) are shown in Figure 52. It should be noted that the blue LED has clearly been diminished in intensity when compared to the green LED by frequency dependent scattering and absorption in the water [33]. The white LED has been diminished by a lesser, but still noticeable amount.

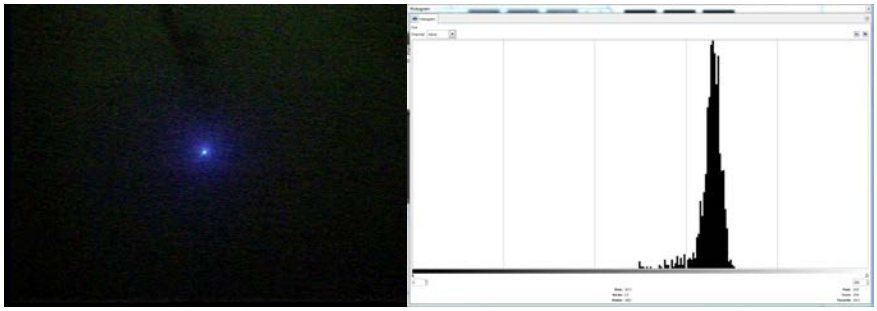
Secondly, it should be noted that despite passing through 13m of turbid sea water, the blue LED still has a very distinct histogram from the green LED. This demonstrates that using coloured LEDs is a feasible method of identifying lights from one another for pose estimation.



(a) *White LED.* ***(b)*** *White LED Hue Histogram.*



(c) *Green LED.* ***(d)*** *Green LED Hue Histogram.*



(e) *Blue LED.* ***(f)*** *Blue LED Hue Histogram.*

Figure 52: *Images of colour LEDs taken at a range of 13m and a depth of 20m.*

7 Conclusion

We have demonstrated an effective means of determining the pose of a submerged UUV using LED lights and a monocular camera. This process is made more difficult by changing ambient light conditions and turbidity in the water. Our system uses a combination of computer-vision algorithms and adapts their parameters during the course of UUV approach to maintain maximum sensitivity at long ranges and noise rejection and accuracy at close range. Pose estimation of the UUV was found using at least three LEDs, but is improved with four or five LEDs visible. The software system was able to run at data rates high enough for feedback control.

These results are ideal for building a computer model of an optical tracking system and its errors and limitations. This is being done for a computer simulation being developed for establishing the viability docking a UUV with a submarine.

There are a number of potential improvements to the methods, particularly in the area of LED identification, but on the whole the system was very effective across a wide variety of UUV positions and orientations. It remains for us to test this system against a live UUV manoeuvering realistically, and as part of an active dock to understand how the movement and dynamics would affect performance.

This page intentionally left blank.

References

[1] Watt, G., Roy, A., Currie, J., Gillis, C., Giesbrecht, J., Heard, G., Birsan, M., Seto, M., Carretero, J., Dubay, R., and Jeans, T. L. (2016), A Concept for Docking a UUV With a Slowly Moving Submarine Under Waves, *IEEE Journal of Oceanic Engineering*, 41(2).

[2] Jaffe, J. S., Moore, K. D., McLean, J., and Strand, M. R. (2001), Underwater optical imaging: Status and prospects, *Oceanography*, 14(3), 64–75.

[3] Giesbrecht, J. and Scrutton, J. (2017), Underwater Optical Imaging to Aid in Docking An Unmanned Underwater Vehicle to a Submarine: Preliminary Experimental Investigations, DRDC Scientific Report, DRDC-RDDC-2017-R155. Defence Research and Development Canada.

[4] Lee, K. W., Kalyan, B., Wijesoma, S., Adams, M., Hover, F., and Patrikalakis, N. (2010), Tracking Random Finite Objects using 3D-LIDAR in Marine Environments, In *Proceedings of 2010 ACM Symposium on Applied Computing*.

[5] Dill, T. (2014), Modeling the Performance of a Laser for Tracking an Underwater Dynamic Target, Ph.D. thesis, Florida Atlantic University.

[6] McLeod, D. and Jacobson, J. (2013), Autonomous Inspection using an Underwater 3D Lidar, In *Proceedings of IEEE Oceans*.

[7] Sattar, J. and Dudek, G. (2006), On the Performance of Color Tracking Algorithms for Underwater Robots under Varying Lighting and Visibility, In *Proceedings of the IEEE Intl. Conference on Robotics and Automation*.

[8] Drews, P., Botelho, S., and Gomes, S. (2008), SLAM in Underwater Environment using SIFT and Topologic Maps, In *Proceeedings IEEE Latin American Robotics Symposium*, pp. 91–96.

[9] Jasiobedzki, P., Se, S., Bondy, M., and Jakola, R. (2008), Underwater 3D mapping and pose estimation for ROV operations, In *Proceedings of IEEE Oceans*.

[10] Garcia, R. and Gracias, N. (2011), Detection of interest points in turbid underwater images, In *Proceedings of IEEE Oceans*.

[11] Massot-Campos, M. and Oliver-Codina, G. (2015), Optical Sensors and Methods for Underwater 3D Reconstruction, *Sensors*, 15(12), 31525-31557.

[12] Cowen, S., Breist, S., and Dombrowski, J. (1997), Underwater Docking of Autonomous Undersea Vehicles Using Optical Terminal Guidance, In *Proceedings IEEE/OES OCEANS Conference*.

[13] Park, J., Jun, B., Lee, P., and Oh, J. (2009), Experiments on vision guided docking of an autonomous underwater vehicle using one camera, *Ocean Engineering*, 36(1), 48–61.

[14] Krupinski, S. (2008), Investigation of autonomous docking strategies for robotic operation on intervention panels, In *Proceedings of IEED/OES OCEANS Conference*.

[15] Hong, Y., Kim, J., Lee, P., Jeon, B., Oh, K., and Oh, J. (2003), Development of the Homing and Docking Algorithm for AUV, In *Proceedings of the Thirteenth Intl. Offshore and Polar Engineering Conference*.

[16] Farr, N., Bowen, A., Ware, J., and Pontbriand, C. (2010), An integrated underwater optical/acoustic communications system, In *Proceedings of IEEE Oceans*.

[17] Bosch, J., Gracias, N., Ridao, P., Istenic, K., and Ribas, D. (2016), Close-Range Tracking of Underwater Vehicles Using Light Beacons, *Sensors*, 16(4), 429.

[18] Li, Y., Jiang, Y., Cao, J., Wang, B., and Li, Y. (2015), AUV docking experiments based on vision positioning using two cameras, *Ocean Engineering*, 110, 163–173.

[19] Kneig, L., Scaramuzza, D., and Siegwart, R. (2011), A Novel Parametrization of the Perspective-Three-Point Problem for a Direct Computation of Absolute Camera Position and Orientation, In *Proceedings of the IEEE Intl. Conf. on Computer Vision and Pattern Recognition*.

[20] Breitenmoser, A., Kneip, L., and Siegwart, R. (2011), A monocular vision-based system for 6D relative robot localization, In *Proceedings IEEE/RSJ International Conf. on Intelligent Robots and Systems*.

[21] Faessler, M., Mueggler, E., Schwabe, K., and Scaramuzza, D. (2014), A monocular pose estimation system based on infrared leds, In *Proceedings of IEEE Intl. Conf. on Robotics and Automation*.

[22] Bradski, G. and Kaehler, A. (2008), Learning OpenCV: Computer Vision with the OpenCV Library, O'Reilly.

[23] Gonzalez, R. and Woods, R. (1992), Digital Image Processing, Addison Wesley.

[24] Feldman, J. (1979), DTNSRDC revised standard submarine equations of motion, (Technical Report SPD-0393-09) David Taylor Naval Ship Research and Development Center.

[25] Fossen, T. (2011), Handbook of marine craft hydrodynamics and motion control, *John Wiley and Sons*.

[26] Fletcher, R. (1980), Practical Methods of Optimization, John Wiley and Sons.

[27] Nocedal, J. and Wright, S. (2006), Numerical Optimization, Springer Series in Operations Research and Financial Engineering, *Springer*.

[28] Giesbrecht, J. and Watt, G. (2017), Optical Position Sensing Challenges in Autonomously Docking a UUV With A Submerged Submarine, In *Proceedings of IEEE OCEANS Conference, Anchorage, Alaska*.

[29] Watt, G. D., Arsenault, R., Lemczyk, T., Phan, D., and Shepeta, V. (2017), Underwater Pose Validation Rig for the DRDC Acoustic Calibration Barge, DRDC Reference Document, DRDC-RDDC-2017-D066, Defence Research and Development Canada.

[30] Shortis, M. (2015), Calibration Techniques for Accurate Measurements by Underwater Camera Systems, *Sensors*, 15(12).

[31] Kunz, C. and Singh, H. (2008), Hemispherical refraction and camera calibration in underwater vision, In *Proceedings of OCEANS*.

[32] Censi, A., Strubel, J., Brandli, C., Delbruck, T., and Scaramuzza, D. (2013), Low-latency localization by active LED markers tracking using a dynamic vision sensor, In *Proceedings of IEEE/RSJ International Conf. on Intelligent Robots and Systems*.

[33] Baiden, G., Bissiri, Y., and Masoti, A. (2009), Paving the way for future underwater omni-directional wireless optical communication systems, *Ocean Engineering*, 36, 633–640.

This page intentionally left blank.

Annex A Analysis One Camera, Multiple LEDs

This Appendix describes how to formulate a well conditioned least squares B matrix for UUV pose (Section 3.5.1) and the linear equations that give UUV location.

A.1 Equation Formulation

Section 3.5.1 establishes the basic relationships between the pixel coordinates p_j , q_j for LED j in a camera image, the LED x_j , y_j , z_j coordinates in UUV body fixed axes, the unknown location of the UUV body fixed axes origin ξ, μ, ν (Figure 13), and the unknown UUV pose ϕ, θ, ψ relative to camera axes. From (2), (3), and (4) these are:

$$\begin{aligned}\left(a_{11}x_j + a_{12}y_j + a_{13}z_j + \xi\right) P_j - a_{21}x_j - a_{22}y_j - a_{23}z_j - \mu = 0\\ \left(a_{11}x_j + a_{12}y_j + a_{13}z_j + \xi\right) Q_j - a_{31}x_j - a_{32}y_j - a_{33}z_j - \nu = 0.\end{aligned} \tag{A1}$$

In an image showing many individual LEDs, any three ($j = 1, 2, 3$ say) can be used to determine the six position and pose unknowns using the six (A1) equations they give rise to. The first step is to use three of the equations to eliminate ξ, μ, ν which appear linearly in (A1). The two $j = 1$ equations give:

$$\begin{aligned}\mu &= a_{11}P_1x_1 + a_{12}P_1y_1 + a_{13}P_1z_1 - a_{21}x_1 - a_{22}y_1 - a_{23}z_1 + \xi P_1\\ \nu &= a_{11}Q_1x_1 + a_{12}Q_1y_1 + a_{13}Q_1z_1 - a_{31}x_1 - a_{32}y_1 - a_{33}z_1 + \xi Q_1\end{aligned} \tag{A2}$$

and after using (A2) to eliminate mu and nu, the final four equations are:

$$\begin{aligned}P_{21}\xi + P_{x21}a_{11} + P_{y21}a_{12} + P_{z21}a_{13} - a_{21}x_{21} - a_{22}y_{21} - a_{23}z_{21} = 0\\ Q_{21}\xi + Q_{x21}a_{11} + Q_{y21}a_{12} + Q_{z21}a_{13} - a_{31}x_{21} - a_{32}y_{21} - a_{33}z_{21} = 0\\ P_{31}\xi + P_{x31}a_{11} + P_{y31}a_{12} + P_{z31}a_{13} - a_{21}x_{31} - a_{22}y_{31} - a_{23}z_{31} = 0\\ Q_{31}\xi + Q_{x31}a_{11} + Q_{y31}a_{12} + Q_{z31}a_{13} - a_{31}x_{31} - a_{32}y_{31} - a_{33}z_{31} = 0\end{aligned} \tag{A3}$$

where:

$$\begin{aligned}k_{21} &= k_2 - k_1, & k_{31} &= k_3 - k_1, & k &= x, y, z, P, Q\\ K_{k21} &= K_2k_2 - K_1k_1, & K_{k31} &= K_3k_3 - K_1k_1, & K &= P, Q.\end{aligned} \tag{A4}$$

To optimally condition the calculations that follow, the (A3) equation used to solve for ξ should be the one with the largest ξ coefficient (the greatest separation between the LEDs) because this coefficient is a divisor throughout the calculations. Therefore, before associating j indices with the three LEDs, the LEDs are sorted so that either $|P_{21}|$ or $|Q_{21}|$ is maximized. Then ξ should be obtained from either the first or second equation in (A3), whichever has the largest ξ coefficient, and substituted into (A2) and the remaining (A3) equations. Doing this for each group of three LEDs increases the robustness of both the least squares solution for pose and linear solution for location that follows. This speeds up the calculations by reducing both

the number of iterations required and the need for Newton method modification.

This sorting process leads to two possible sets of six equations for each group of three LEDs, only one of which is used depending on the magnitudes of P_{21} and Q_{21}. If $|P_{21}| > |Q_{21}|$ then $|P_{21}|$ provides the best conditioning for this group of LEDs and:

$$\begin{bmatrix} P_{21}\xi \\ P_{21}\mu \\ P_{21}\nu \end{bmatrix} = \mathbf{\Xi}_P \cdot \mathbf{V} \qquad \text{and} \qquad \mathbf{B}_P \cdot \mathbf{V} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{A5}$$

where $\mathbf{V}$ is defined in (5) and:

$$\mathbf{\Xi}_P = \begin{bmatrix} -P_{x21} & -P_{y21} & -P_{z21} & x_{21} & y_{21} & z_{21} & 0 & 0 & 0 \\ -P_1P_2x_{21} & -P_1P_2y_{21} & -P_1P_2z_{21} & P_1x_2 - P_2x_1 & P_1y_2 - P_2y_1 & P_1z_2 - P_2z_1 & 0 & 0 & 0 \\ -Q_1P_2x_{21} & -Q_1P_2y_{21} & -Q_1P_2z_{21} & Q_1x_{21} & Q_1y_{21} & Q_1z_{21} & -x_1P_{21} & -y_1P_{21} & -z_1P_{21} \end{bmatrix} \tag{A6}$$

$$\mathbf{B}_P = \begin{bmatrix} P_{21}Q_{x21} - P_{x21}Q_{21} & P_{21}Q_{y21} - P_{y21}Q_{21} & P_{21}Q_{z21} - P_{z21}Q_{21} & Q_{21}x_{21} & Q_{21}y_{21} & Q_{21}z_{21} & -P_{21}x_{21} & -P_{21}y_{21} & -P_{21}z_{21} \\ P_{21}Q_{x31} - P_{x21}Q_{31} & P_{21}Q_{y31} - P_{y21}Q_{31} & P_{21}Q_{z31} - P_{z21}Q_{31} & Q_{31}x_{21} & Q_{31}y_{21} & Q_{31}z_{21} & -P_{21}x_{31} & -P_{21}y_{31} & -P_{21}z_{31} \\ P_{21}P_{x31} - P_{x21}P_{31} & P_{21}P_{y31} - P_{y21}P_{31} & P_{21}P_{z31} - P_{z21}P_{31} & P_{31}x_{21} - P_{21}x_{31} & P_{31}y_{21} - P_{21}y_{31} & P_{31}z_{21} - P_{21}z_{31} & 0 & 0 & 0 \end{bmatrix}. \tag{A7}$$

Alternatively, if $|P_{21}| < |Q_{21}|$ then $|Q_{21}|$ provides the best group conditioning and the six equations to use are:

$$\begin{bmatrix} Q_{21}\xi \\ Q_{21}\mu \\ Q_{21}\nu \end{bmatrix} = \mathbf{\Xi}_Q \cdot \mathbf{V} \qquad \text{and} \qquad \mathbf{B}_Q \cdot \mathbf{V} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{A8}$$

where:

$$\mathbf{\Xi}_Q = \begin{bmatrix} -Q_{x21} & -Q_{y21} & -Q_{z21} & 0 & 0 & 0 & x_{21} & y_{21} & z_{21} \\ -P_1Q_2x_{21} & -P_1Q_2y_{21} & -P_1Q_2z_{21} & -x_1Q_{21} & -y_1Q_{21} & -z_1Q_{21} & P_1x_{21} & P_1y_{21} & P_1z_{21} \\ -Q_1Q_2x_{21} & -Q_1Q_2y_{21} & -Q_1Q_2z_{21} & 0 & 0 & 0 & Q_1x_2 - Q_2x_1 & Q_1y_2 - Q_2y_1 & Q_1z_2 - Q_2z_1 \end{bmatrix} \tag{A9}$$

$$\mathbf{B}_Q = \begin{bmatrix} Q_{21}P_{x21} - Q_{x21}P_{21} & Q_{21}P_{y21} - Q_{y21}P_{21} & Q_{21}P_{z21} - Q_{z21}P_{21} & -Q_{21}x_{21} & -Q_{21}y_{21} & -Q_{21}z_{21} & P_{21}x_{21} & P_{21}y_{21} & P_{21}z_{21} \\ Q_{21}P_{x31} - Q_{x21}P_{31} & Q_{21}P_{y31} - Q_{y21}P_{31} & Q_{21}P_{z31} - Q_{z21}P_{31} & -Q_{21}x_{31} & -Q_{21}y_{31} & -Q_{21}z_{31} & P_{31}x_{21} & P_{31}y_{21} & P_{31}z_{21} \\ Q_{21}Q_{x31} - Q_{x21}Q_{31} & Q_{21}Q_{y31} - Q_{y21}Q_{31} & Q_{21}Q_{z31} - Q_{z21}Q_{31} & 0 & 0 & 0 & Q_{31}x_{21} - Q_{21}x_{31} & Q_{31}y_{21} - Q_{21}y_{31} & Q_{31}z_{21} - Q_{21}z_{31} \end{bmatrix}. \tag{A10}$$

Note that the condition measure, P_{21}, for (A5) is left to multiply by the left hand side of all size of these equations. This provides condition-based weighted averaging of the contributions of each three LED group to the overall result.

The least squares formulation of the UUV pose problem, and the linear solution to the UUV localization problem, are obtained by combining the equations from all *N* groups of the three LEDs:

$$\begin{aligned} \mathbf{\Xi} &= \sum \operatorname{sign}(K_{21})\, \mathbf{\Xi}_K \\ \sigma &= \sum |K_{21}| \\ \mathbf{B} &= \text{concatenate-by-row}(\mathbf{B}_K) \end{aligned} \tag{A11}$$

where K is the local conditioning mode, either P or Q, and $\mathbf{B}$ is the $3N \times 9$ matrix used in (6). We can now pursue a least squares solution to $\mathbf{B} \cdot \mathbf{V} = \mathbf{0}$ to get the pose ϕ, θ, ψ, as described in 3.5.1 beginning with (6). This allows $\mathbf{V}$ to be evaluated. Then:

$$\begin{bmatrix} \xi \\ \mu \\ \nu \end{bmatrix} = \frac{\mathbf{\Xi}}{\sigma} \cdot \mathbf{V}. \tag{A12}$$

A.2 Gradient and Hessian Construction

To minimize the sum of the squares $S = \mathbf{V}^T \cdot \mathbf{B}^T\mathbf{B} \cdot \mathbf{V}$ from (6), we must calculate the gradient $\mathbf{G}$ and Hessian $\mathbf{H}$ from (7). Since $\mathbf{B}^T\mathbf{B}$ is symmetric and independent of the Euler angles (represented here by χ):

$$\frac{\partial S}{\partial \chi} = \frac{\partial \mathbf{V}^T}{\partial \chi} \cdot \mathbf{B}^T\mathbf{B} \cdot \mathbf{V} + \mathbf{V} \cdot \mathbf{B}^T\mathbf{B} \cdot \frac{\partial \mathbf{V}}{\partial \chi} = 2\frac{\partial \mathbf{V}^T}{\partial \chi} \cdot \mathbf{B}^T\mathbf{B} \cdot \mathbf{V} \tag{A13}$$

so that:

$$\mathbf{G} = \begin{bmatrix} \partial S/\partial \phi \\ \partial S/\partial \theta \\ \partial S/\partial \psi \end{bmatrix} = 2 \begin{bmatrix} \partial \mathbf{V}^T/\partial \phi \\ \partial \mathbf{V}^T/\partial \theta \\ \partial \mathbf{V}^T/\partial \psi \end{bmatrix} \cdot \mathbf{B}^T\mathbf{B} \cdot \mathbf{V}. \tag{A14}$$

where:

$$\frac{\partial \mathbf{V}}{\partial \phi} = \begin{bmatrix} 0 \\ a_{13} \\ -a_{12} \\ 0 \\ a_{23} \\ -a_{22} \\ 0 \\ a_{33} \\ -a_{32} \end{bmatrix}, \quad \frac{\partial \mathbf{V}}{\partial \theta} = \begin{bmatrix} -\sin\theta\cos\psi \\ \sin\phi\, a_{11} \\ \cos\phi\, a_{11} \\ -\sin\theta\sin\psi \\ \sin\phi\, a_{21} \\ \cos\phi\, a_{21} \\ -\cos\theta \\ -\sin\phi\sin\theta \\ -\cos\phi\sin\theta \end{bmatrix}, \quad \frac{\partial \mathbf{V}}{\partial \psi} = \begin{bmatrix} -a_{21} \\ -a_{22} \\ -a_{23} \\ a_{11} \\ a_{12} \\ a_{13} \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{A15}$$

The Hessian is symmetric:

$$\mathbf{H} = \begin{bmatrix} \partial^2 S/\partial\phi^2 & \partial^2 S/\partial\phi\partial\theta & \partial^2 S/\partial\phi\partial\psi \\ \partial^2 S/\partial\theta\partial\phi & \partial^2 S/\partial\theta^2 & \partial^2 S/\partial\theta\partial\psi \\ \partial^2 S/\partial\psi\partial\phi & \partial^2 S/\partial\psi\partial\theta & \partial^2 S/\partial\psi^2 \end{bmatrix}.$$

Its diagonal terms are:

$$\frac{\partial^2 S}{\partial \chi^2} = 2\frac{\partial^2 \mathbf{V}}{\partial \chi^2} \cdot \mathbf{B}^T\mathbf{B} \cdot \mathbf{V} + 2\frac{\partial \mathbf{V}}{\partial \chi} \cdot \mathbf{B}^T\mathbf{B} \cdot \frac{\partial \mathbf{V}}{\partial \chi} \tag{A16}$$

where:

$$\frac{\partial^2 \mathbf{V}}{\partial \phi^2} = \begin{bmatrix} 0 \\ -a_{12} \\ -a_{13} \\ 0 \\ -a_{22} \\ -a_{23} \\ 0 \\ -a_{32} \\ -a_{33} \end{bmatrix}, \quad \frac{\partial^2 \mathbf{V}}{\partial \theta^2} = \begin{bmatrix} -a_{11} \\ -\sin\phi\sin\theta\cos\psi \\ -\cos\phi\sin\theta\cos\psi \\ -a_{21} \\ -\sin\phi\sin\theta\sin\psi \\ -\cos\phi\sin\theta\sin\psi \\ \sin\theta \\ -a_{32} \\ -a_{33} \end{bmatrix}, \quad \frac{\partial^2 \mathbf{V}}{\partial \psi^2} = \begin{bmatrix} -a_{11} \\ -a_{12} \\ -a_{13} \\ -a_{21} \\ -a_{22} \\ -a_{23} \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{A17}$$

The off-diagonal Hessian terms are, for example:

$$\frac{\partial^2 S}{\partial\phi\partial\theta} = 2\frac{\partial^2 \mathbf{V}^T}{\partial\phi\partial\theta}\cdot\mathbf{B}^T\mathbf{B}\cdot\mathbf{V} + 2\frac{\partial \mathbf{V}^T}{\partial\phi}\cdot\mathbf{B}^T\mathbf{B}\cdot\frac{\partial\mathbf{V}}{\partial\theta} \tag{A18}$$

and these can be evaluated using (A15) and:

$$\frac{\partial^2\mathbf{V}}{\partial\phi\partial\theta} = \begin{bmatrix} 0 \\ \partial\mathbf{V}/\partial\theta|_3 \\ -\partial\mathbf{V}/\partial\theta|_2 \\ 0 \\ \partial\mathbf{V}/\partial\theta|_6 \\ -\partial\mathbf{V}/\partial\theta|_5 \\ 0 \\ \partial\mathbf{V}/\partial\theta|_9 \\ -\partial\mathbf{V}/\partial\theta|_8 \end{bmatrix}, \quad \frac{\partial^2\mathbf{V}}{\partial\phi\partial\psi} = \begin{bmatrix} 0 \\ -a_{23} \\ a_{22} \\ 0 \\ a_{13} \\ -a_{12} \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \frac{\partial^2\mathbf{V}}{\partial\theta\partial\psi} = \begin{bmatrix} -\partial\mathbf{V}/\partial\theta|_4 \\ -\partial\mathbf{V}/\partial\theta|_5 \\ -\partial\mathbf{V}/\partial\theta|_6 \\ \partial\mathbf{V}/\partial\theta|_1 \\ \partial\mathbf{V}/\partial\theta|_2 \\ \partial\mathbf{V}/\partial\theta|_3 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{A19}$$

This page intentionally left blank.

Annex B Analysis Multiple Cameras, Multiple LEDs

This Annex considers the changes that need to be made to the "one camera, multiple LEDs" analysis when multiple cameras are used. Unlike the analysis of Annex A, which has been numerically implemented, verified, and validated, the analysis here is theoretical and presented in anticipation of a future need.

If there are m cameras in Figure 13, then the "camera" frame coordinates need indices, as in ξ^i, μ^i, ν^i, for $i = 1 \ldots m$. As well, the orientation of the UUV relative to camera i is now denoted by ϕ^i, θ^i, ψ^i and, from (2), the associated Euler angle transformation is $\mathbf{A}^i = \mathbf{A}(\phi^i, \theta^i, \psi^i)$.

As before, our objective is to find the UUV position ξ, μ, ν and orientation ϕ, θ, ψ relative to a known frame of reference which we denote the "common" frame of reference. All m camera positions and orientations must be known relative to the common frame. For generality, we assume each camera is mounted on a base and that the camera position and orientation can be adjusted relative to that base. This adjustment might be permanent or a function of time (if the camera can be panned, for example). So there needs to be an intermediate "base" coordinate system between the common and each camera's reference frame.

Let $\xi^{bi}, \mu^{bi}, \nu^{bi}$ be the base coordinate system for camera i. Its location and pose relative to the common frame are $\xi_{bi}, \mu_{bi}, \nu_{bi}, \phi_{bi}, \theta_{bi}, \psi_{bi}$. The camera i frame ξ^i, μ^i, ν^i is located and oriented at $\xi^{bi}, \mu^{bi}, \nu^{bi}, \phi^{bi}, \theta^{bi}, \psi^{bi}$ relative to the camera i base frame.

With this notation, the location of LED j in the common frame is, analogously to (4):

$$\begin{aligned}
\begin{bmatrix} \xi_j \\ \mu_j \\ \nu_j \end{bmatrix} &= \begin{bmatrix} \xi \\ \mu \\ \nu \end{bmatrix} + \mathbf{A} \begin{bmatrix} x_j \\ y_j \\ z_j \end{bmatrix} \\
&= \begin{bmatrix} \xi_{bi} \\ \mu_{bi} \\ \nu_{bi} \end{bmatrix} + \mathbf{A}_{bi} \left\{ \begin{bmatrix} \xi^{bi} \\ \mu^{bi} \\ \nu^{bi} \end{bmatrix} + \mathbf{A}^{bi} \left(\begin{bmatrix} \xi^i \\ \mu^i \\ \nu^i \end{bmatrix} + \mathbf{A}^i \begin{bmatrix} x_j \\ y_j \\ z_j \end{bmatrix} \right) \right\} \qquad \text{(B1)} \\
&= \begin{bmatrix} \xi_{ci} \\ \mu_{ci} \\ \nu_{ci} \end{bmatrix} + \mathbf{A}_{bi} \cdot \mathbf{A}^{bi} \begin{bmatrix} \xi^i \\ \mu^i \\ \nu^i \end{bmatrix} + \mathbf{A}_{bi} \cdot \mathbf{A}^{bi} \cdot \mathbf{A}^i \begin{bmatrix} x_j \\ y_j \\ z_j \end{bmatrix}
\end{aligned}$$

where $\mathbf{A}_{bi} = \mathbf{A}(\phi_{bi}, \theta_{bi}, \psi_{bi})$, $\mathbf{A}^{bi} = \mathbf{A}(\phi^{bi}, \theta^{bi}, \psi^{bi})$, and:

$$\begin{bmatrix} \xi_{ci} \\ \mu_{ci} \\ \nu_{ci} \end{bmatrix} = \begin{bmatrix} \xi_{bi} \\ \mu_{bi} \\ \nu_{bi} \end{bmatrix} + \mathbf{A}_{bi} \begin{bmatrix} \xi^{bi} \\ \mu^{bi} \\ \nu^{bi} \end{bmatrix}. \qquad \text{(B2)}$$

This results in:

$$\begin{bmatrix} \xi^i \\ \mu^i \\ \nu^i \end{bmatrix} = \mathbf{Z}^i \begin{bmatrix} \bar{\xi} \\ \bar{\mu} \\ \bar{\nu} \end{bmatrix} \qquad \text{and} \qquad \mathbf{A}^i = \mathbf{Z}^i \cdot \mathbf{A} \tag{B3}$$

where:

$$\mathbf{Z}^i = (\mathbf{A}^{bi})^{-1} \cdot \mathbf{A}_{bi}^{-1}, \quad \bar{\xi} = \xi - \xi_{ci}, \quad \bar{\mu} = \mu - \mu_{ci}, \quad \bar{\nu} = \nu - \nu_{ci} \tag{B4}$$

Since the $\mathbf{A}$ matrices (and $\mathbf{Z}^i$) are all orthogonal, the inverse of each is equal to its transpose. If $\mathbf{V}^i$ is created from $\mathbf{A}^i$ as $\mathbf{V}$ was from $\mathbf{A}$, then $\mathbf{Z}^i$ gives the relationship between the two:

$$\mathbf{V}^i = \mathbf{Z}_V^i \cdot \mathbf{V} \tag{B5}$$

where, if the components of $\mathbf{Z}^i$ are $z_{11}^i, z_{12}^i, \ldots$, then:

$$\mathbf{Z}_V^i \equiv \begin{bmatrix} z_{11}^i & 0 & 0 & z_{12}^i & 0 & 0 & z_{13}^i & 0 & 0 \\ 0 & z_{11}^i & 0 & 0 & z_{12}^i & 0 & 0 & z_{13}^i & 0 \\ 0 & 0 & z_{11}^i & 0 & 0 & z_{12}^i & 0 & 0 & z_{13}^i \\ z_{21}^i & 0 & 0 & z_{22}^i & 0 & 0 & z_{23}^i & 0 & 0 \\ 0 & z_{21}^i & 0 & 0 & z_{22}^i & 0 & 0 & z_{23}^i & 0 \\ 0 & 0 & z_{21}^i & 0 & 0 & z_{22}^i & 0 & 0 & z_{23}^i \\ z_{31}^i & 0 & 0 & z_{32}^i & 0 & 0 & z_{33}^i & 0 & 0 \\ 0 & z_{31}^i & 0 & 0 & z_{32}^i & 0 & 0 & z_{33}^i & 0 \\ 0 & 0 & z_{31}^i & 0 & 0 & z_{32}^i & 0 & 0 & z_{33}^i \end{bmatrix}. \tag{B6}$$

Now, from (A1), the image taken by camera i of LED j results in two equations:

$$\mathbf{M}_j^i \cdot \mathbf{V}^i + \mathbf{N}_j^i \begin{bmatrix} \xi^i \\ \mu^i \\ \nu^i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{B7}$$

where:

$$\begin{aligned} \mathbf{M}_j^i &= \begin{bmatrix} x_j P_j^i & y_j P_j^i & z_j P_j^i & -x_j & -y_j & -z_j & 0 & 0 & 0 \\ x_j Q_j^i & y_j Q_j^i & z_j Q_j^i & 0 & 0 & 0 & -x_j & -y_j & -z_j \end{bmatrix} \\ \mathbf{N}_j^i &= \begin{bmatrix} P_j^i & -1 & 0 \\ Q_j^i & 0 & -1 \end{bmatrix}. \end{aligned} \tag{B8}$$

These equations can be reformulated in terms of the six unknown common frame variables $\xi, \mu, \nu, \phi, \theta, \psi$ with all other parameters known:

$$\mathbf{M}_j^i \cdot \mathbf{Z}_V^i \cdot \mathbf{V} + \mathbf{N}_j^i \cdot \mathbf{Z}^i \begin{bmatrix} \bar{\xi} \\ \bar{\mu} \\ \bar{\nu} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{B9}$$

With the following definitions:

$$P^i_{jzk} \equiv P^i_j z^i_{1k} - z^i_{2k} \qquad \text{and} \qquad Q^i_{jzk} \equiv Q^i_j z^i_{1k} - z^i_{3k} \tag{B10}$$

(B9) can be written:

$$\begin{bmatrix} P^i_{jz1} & P^i_{jz2} & P^i_{jz3} \\ Q^i_{jz1} & Q^i_{jz2} & Q^i_{jz3} \end{bmatrix} \left(\mathbf{A} \begin{bmatrix} x_j \\ y_j \\ z_j \end{bmatrix} + \begin{bmatrix} \bar{\xi} \\ \bar{\mu} \\ \bar{\nu} \end{bmatrix} \right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{B11}$$

which provides a convenient mechanism for generating the necessary coefficients.

Each camera i and LED j combination gives rise to two such equations. With three unique i, j combinations, we have the necessary six equations to solve for the common frame unknowns. Ideally, however, we will have more than just three combinations as this will increase the robustness of the solution; the more combinations (equations) the better. These equations can be solved the same way as before, by taking all possible i, j combinations three at a time and building up condition-weighted coefficients for the unknowns.

The six equations resulting from each group of three combinations are all linear in $\bar{\xi}, \bar{\mu}, \bar{\nu}$. Again, it is desirable to eliminate these three unknowns and reduce the number of equations to three before carrying out the nonlinear solution of the Euler angles. To facilitate this, consider the linear portions of any three of the equations:

$$\begin{bmatrix} K^1_{z1} & K^1_{z2} & K^1_{z3} \\ K^2_{z1} & K^2_{z2} & K^2_{z3} \\ K^3_{z1} & K^3_{z2} & K^3_{z3} \end{bmatrix} \begin{bmatrix} \bar{\xi} \\ \bar{\mu} \\ \bar{\nu} \end{bmatrix} \equiv \mathbf{K} \begin{bmatrix} \bar{\xi} \\ \bar{\mu} \\ \bar{\nu} \end{bmatrix} \tag{B12}$$

where $K = P$ or Q and different K superscripts indicate unique P, Q, i, j combinations; it is possible, for example, for the i, j combination to be the same as long as K is different (as in (A2), for example).

Expressions for the $\bar{\xi}, \bar{\mu}, \bar{\nu}$ unknowns in each group of six equations can be found by inverting the (B12) $\mathbf{K}$ coefficient matrix and multiplying the three equations from which $\mathbf{K}$ was obtained by $\mathbf{K}^{-1}$. However, this will only work if $\mathbf{K}$ is not singular, and it will work best if $\mathbf{K}$ is not close to being singular. The $\mathbf{K}$ matrix becomes singular as the determinant of $\mathbf{K}$, denoted $\det(\mathbf{K})$, goes to zero. From our six equations, there are 20 possible groupings of three equations to choose $\mathbf{K}$ from. The best group will be the one with the largest $\det(\mathbf{K})$ magnitude. To find this group, we must cycle through all 20 possibilities calculating $\det(\mathbf{K})$ for each one, which is not difficult. This process is equivalent to that used in Section A.1 except, there, the μ and ν equations (A2) are restricted to the same LED, which has no impact on $\det(\mathbf{K})$ in that special case.

Having identified the three equations that maximize the magnitude of $\det(\mathbf{K})$, we use:

$$w \equiv |\det(\mathbf{K})| \tag{B13}$$

as the condition-weight for this group of six equations. That is, we solve for $w\bar{\xi}, w\bar{\mu}, w\bar{\nu}$ and use the coefficients of the $\mathbf{A}$ elements from these equations to create $\mathbf{\Xi}_K$, analogously to (A6) and (A9). These $\bar{\xi}, \bar{\mu}, \bar{\nu}$ expressions are also substituted into the remaining three group equations which are themselves condition-weighted by multiplying them by w. These latter equations will now be composed of only linear terms in the $\mathbf{A}$ elements and their coefficients become the elements of a $\mathbf{B}_K$ matrix, analogously to (A7) and A(10).

This process is repeated for each unique group of three i, j combinations while building up $\mathbf{\Xi}$, $\sigma = \sum w$, and $\mathbf{B}$ as per (A11). The solution of the six unknowns can then proceed as before.

Annex C Underwater Camera Optics

This Annex models the optical characteristics of the SubC underwater camera discussed in Section 4.1. The model forms the basis of the analysis presented in Section 3.5.1.

Figure C.1 provides a schematic of the camera viewing an underwater object, and shows how the object is perceived by the camera's charge coupled device (CCD).

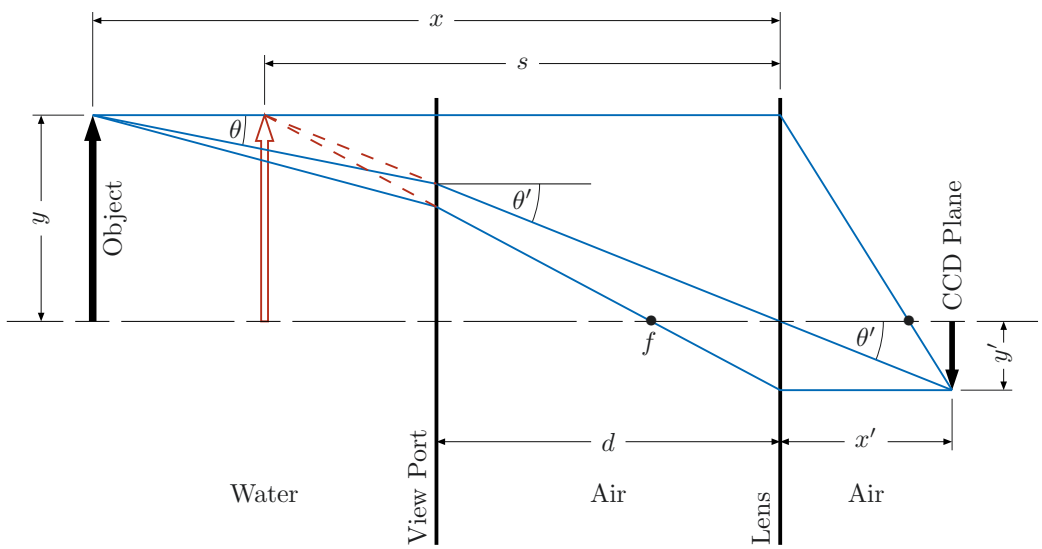


Figure C.1: *Schematic showing how the SubC underwater camera perceives an underwater object. Watt et al [29] give* $d = 45.9$ *mm and* $x' = 16.8$ *mm.*

The camera is secured in a waterproof housing with its lens a distance d behind the viewing port in the housing. There is air inside the housing and the camera. Light from an object in the water is refracted through the flat sided viewing port according to Snell's law, which models light passing through an interface between two media as:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2$$

where n is the index of refraction for a medium and θ is the angle a light ray makes with a normal to the interface. Light passing through the glass viewing port sees two interfaces, a water/glass followed by a glass/air interface, which results in:

$$n \sin\theta = n_g \sin\theta_g = \sin\theta' \tag{C1}$$

where $n, n_g, 1$ are the indices of refraction for seawater, glass, and air, and $\theta, \theta_g, \theta'$ are the angles a light ray makes relative to the parallel interface normals in each medium. This shows, given n and θ, that the glass has no impact on θ'. However, the glass does add an increment to x that is approximately equal to the glass thickness, which we ignore.

The camera has a horizontal field of view of 55 degrees, so $\max(\theta') = 27.5$ degrees and $\max(\theta) = \sin^{-1}(\sin(27.5)/n) = 20.3$ degrees assuming $n = 1.33$. The vertical field of view is 42.5 degrees so the maximum θ' and θ magnitudes in the vertical plane are 21.25 and 15.8 degrees, respectively.

Consider the light ray in Figure C.1 that passes through the lens at the camera centreline. Trigonometry gives:

$$\tan\theta' = \frac{y'}{x'} = \frac{y}{s} \tag{C2}$$

and:

$$(s-d)\tan\theta' = (x-d)\tan\theta. \tag{C3}$$

Eliminating θ from this latter expression using (A1) gives:

$$\frac{s-d}{x-d} = \frac{\tan\theta}{\tan\theta'} = \frac{1}{\sqrt{(n^2-1)\tan^2\theta' + n^2}} \tag{C4}$$

Solving for s, (C2) becomes:

$$\begin{aligned}\tan\theta' = & \frac{y}{d + \dfrac{(x-d)}{\sqrt{(n^2-1)\tan^2\theta' + n^2}}} \\ & \sim \frac{ny}{x+(n-1)d}\left[1 + \frac{(x-d)}{2(x+(n-1)d)}\left(\frac{(n^2-1)\tan^2\theta'}{n^2}\right)\right. \\ & \left. + O\left(\frac{(n^2-1)\tan^2\theta'}{n^2}\right)^2\right]\end{aligned} \tag{C5}$$

In Section 3.5.1, we assume:

$$\frac{y'}{nx'} = \frac{y}{x+x'} \tag{C6}$$

This ignores the $O(\tan^2\theta')$ term in (C5) which, for $n = 1.33$, the worst case situation of $\theta' = 27.5$ degrees, and assuming $d \ll x$, gives a worst case error of 6%. The average error is less than this because the average $|\theta'|$ is less than 27.5 degrees. Also, nx' is treated as an empirically determined calibration constant and optimized in units of pixels using the entire data set. This is necessary because y' is only known in pixel coordinates, not as an actual length (the physical dimensions of the CCD array are not precisely known).

Annex D Position and Pose Results For All Tests

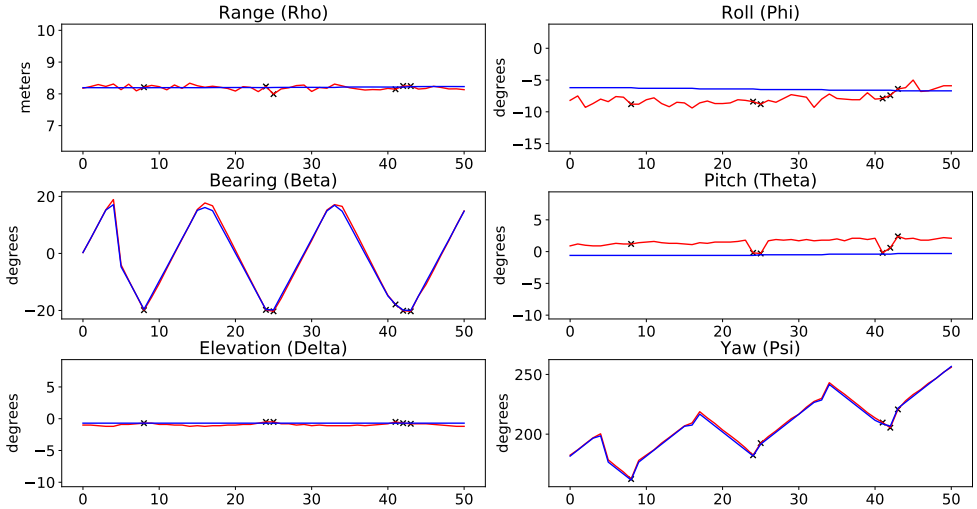


Figure D.1: *8m range, 20m depth, no elevation offset (Test 6).*

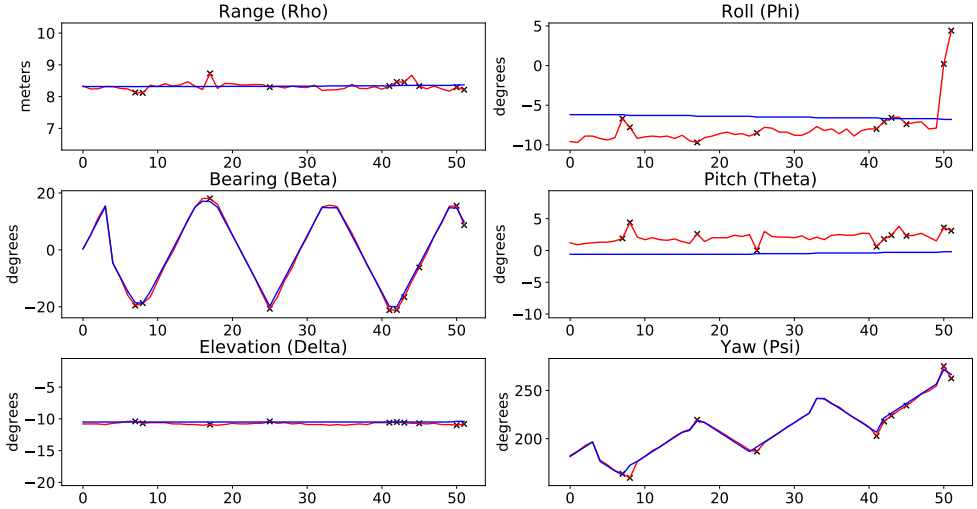


Figure D.2: *8m range, 20m depth, 1.4m elevation offset (Test 8).*

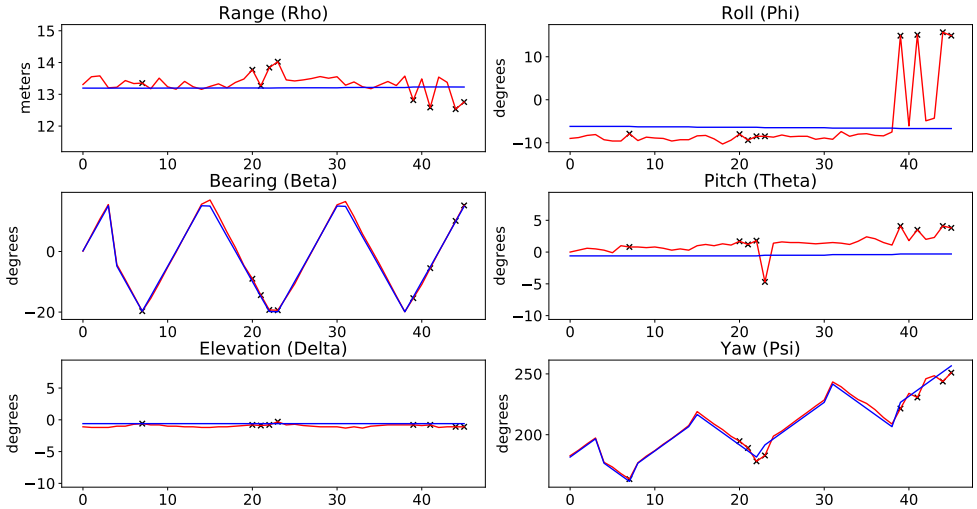


Figure D.3: *13m range, 20m depth, no elevation offset (Test 9).*

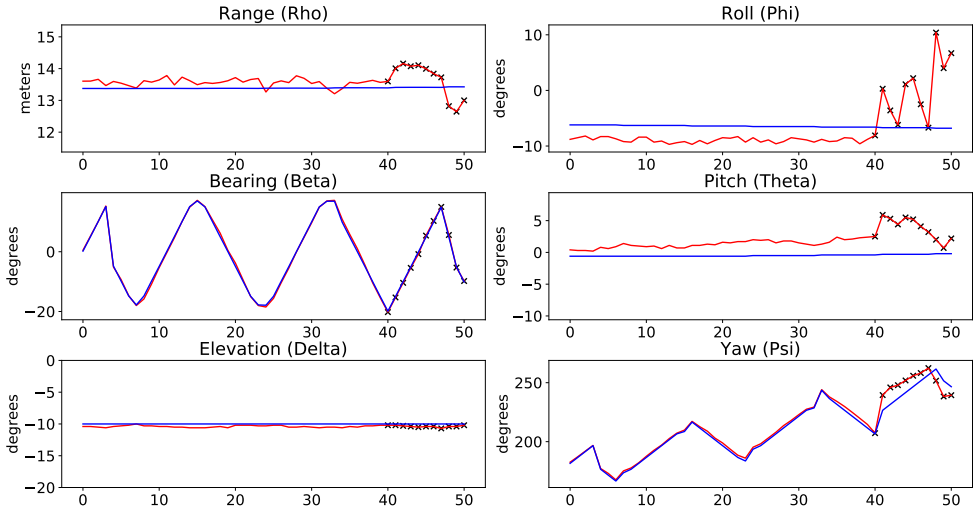


Figure D.4: *13m range, 20m depth, 2.2m elevation offset (Test 11).*

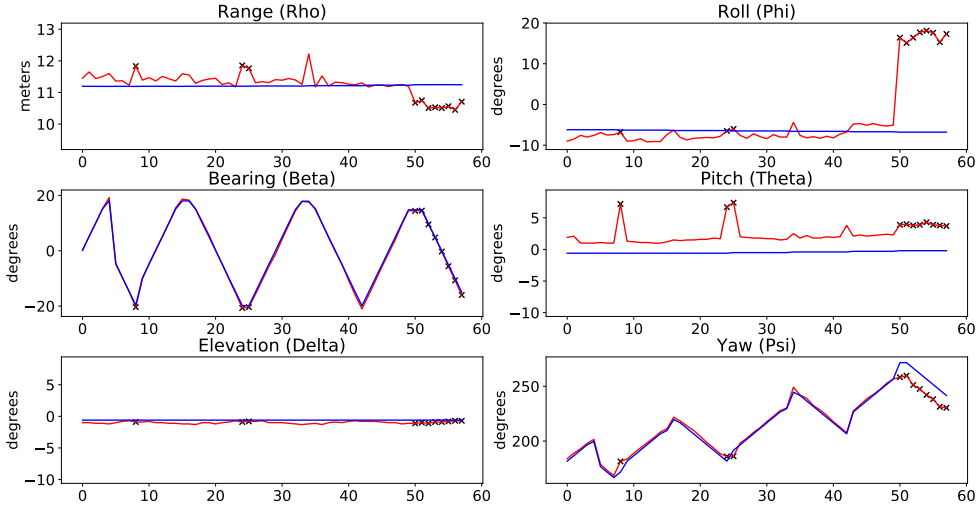


Figure D.5: *11m range, 20m depth, no elevation offset (Test 13).*

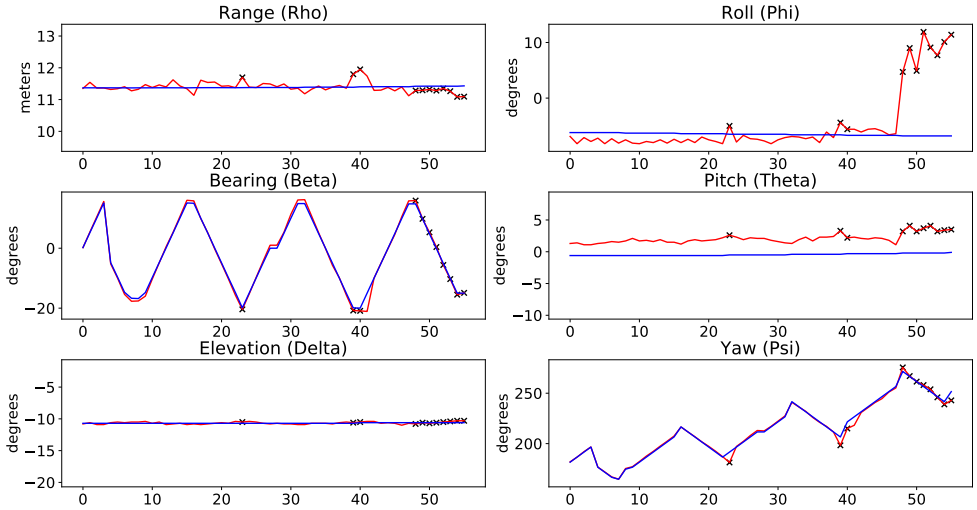


Figure D.6: *11m range, 20m depth, 2.0m elevation offset (Test 14).*

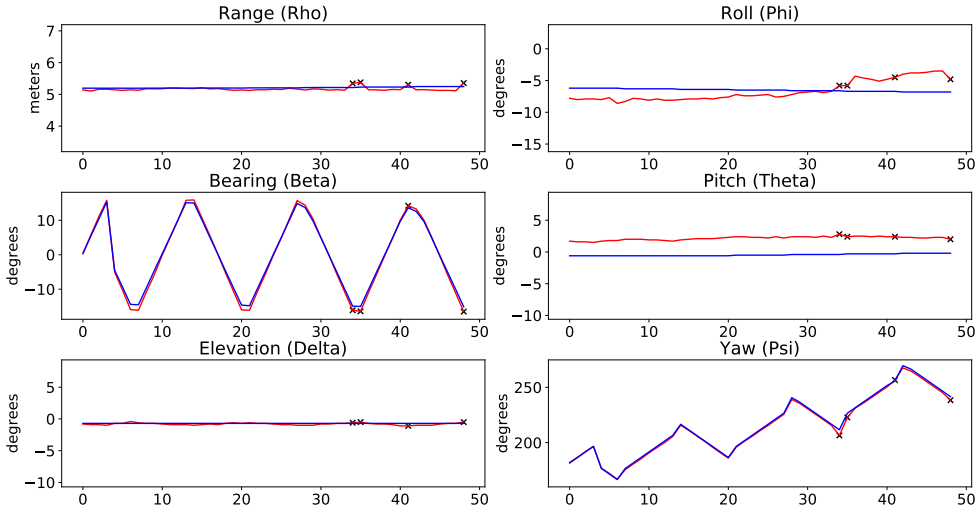


Figure D.7: *5m range, 20m depth, no elevation offset (Test 17).*

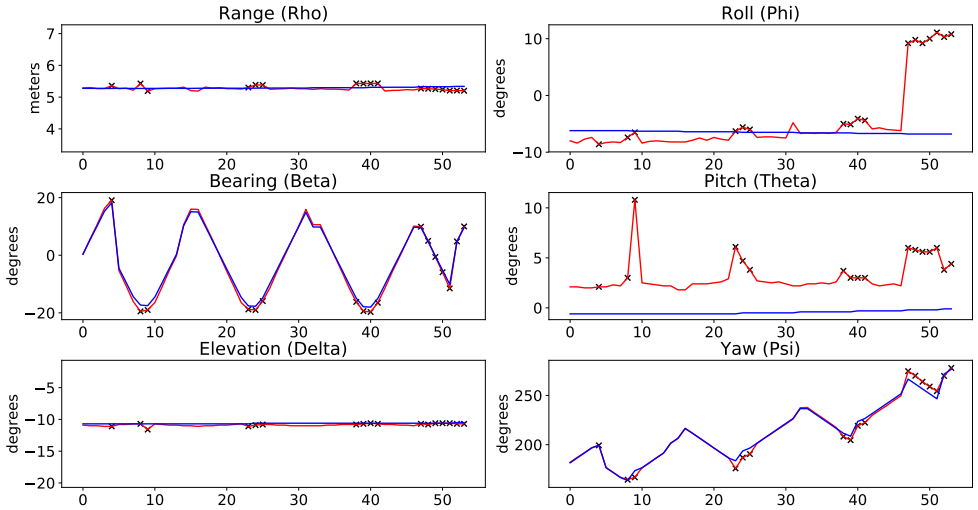


Figure D.8: *5m range, 20m depth, 0.9m elevation offset (Test 18).*

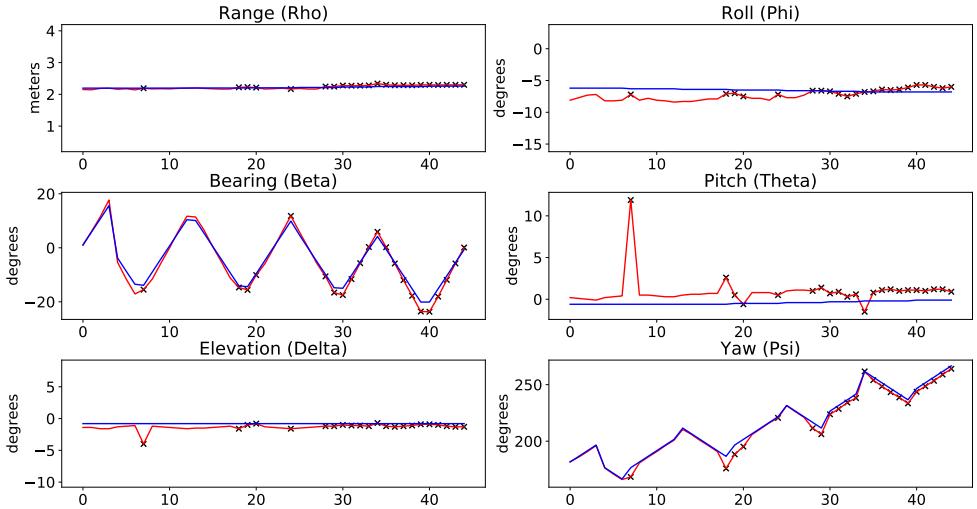


Figure D.9: *2m range, 20m depth, no elevation offset (Test 19).*

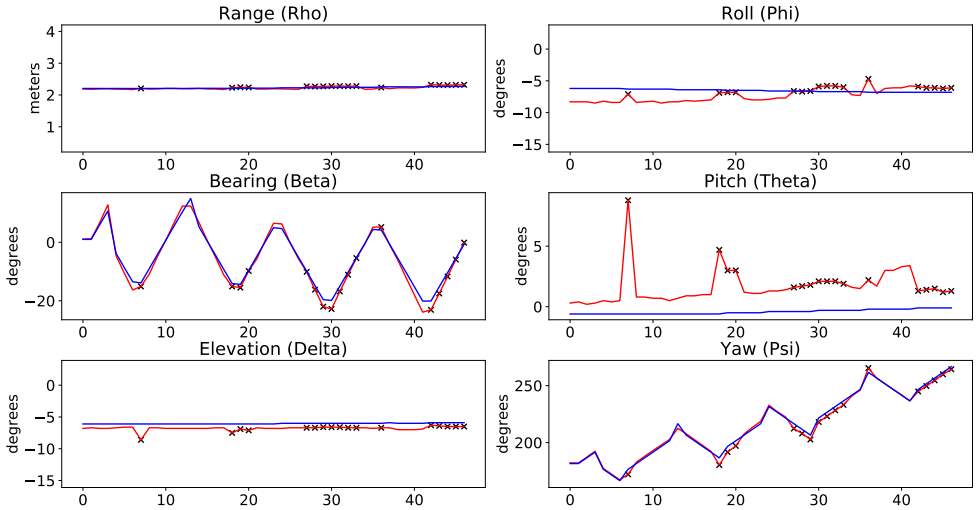


Figure D.10: *2m range, 20m depth, 0.2m elevation offset (Test 22).*

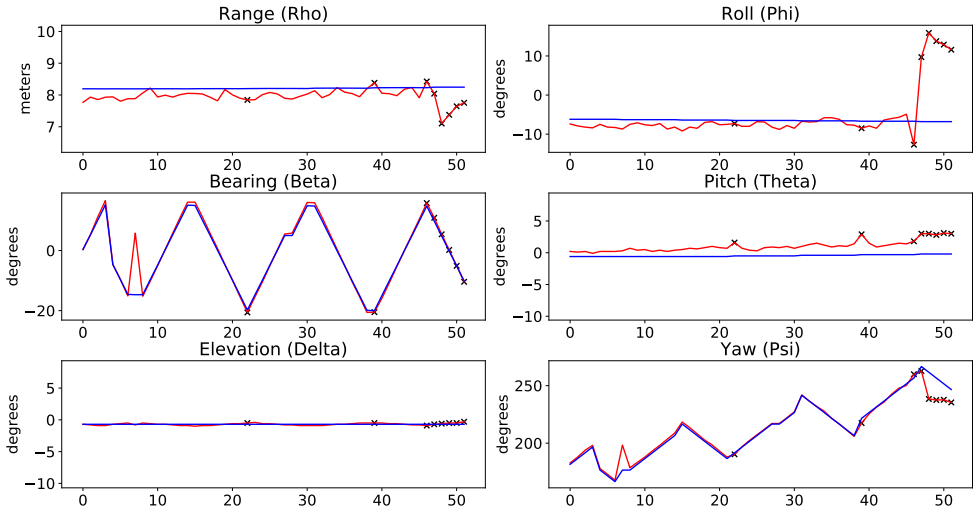


Figure D.11: *8m range, 5m depth, no elevation offset (Test 24).*

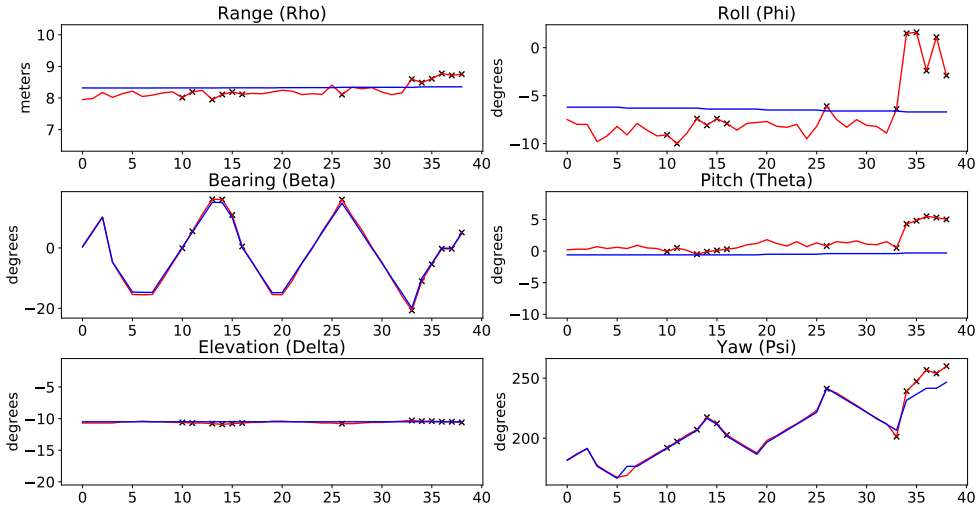


Figure D.12: *8m range, 5m depth, 1.4m elevation offset (Test 25).*

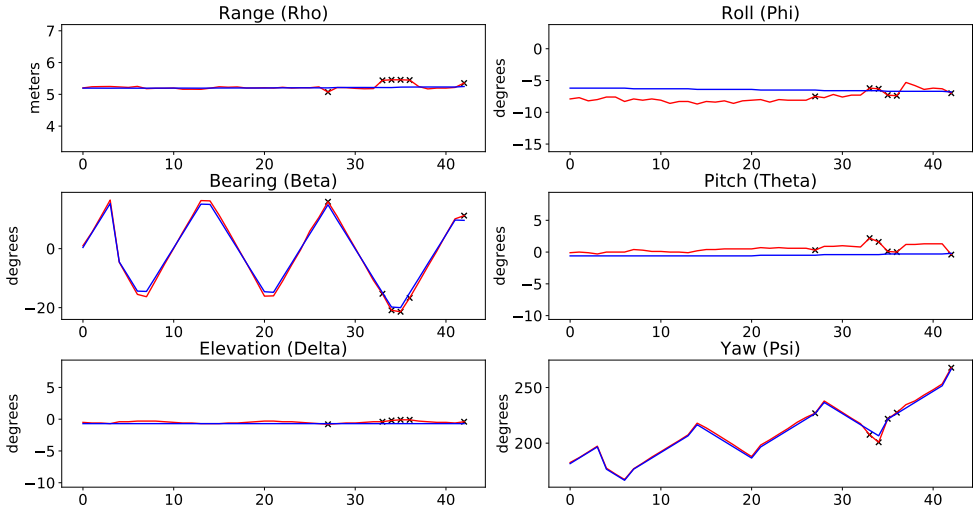


Figure D.13: *5m range, 5m depth, no elevation offset (Test 26).*

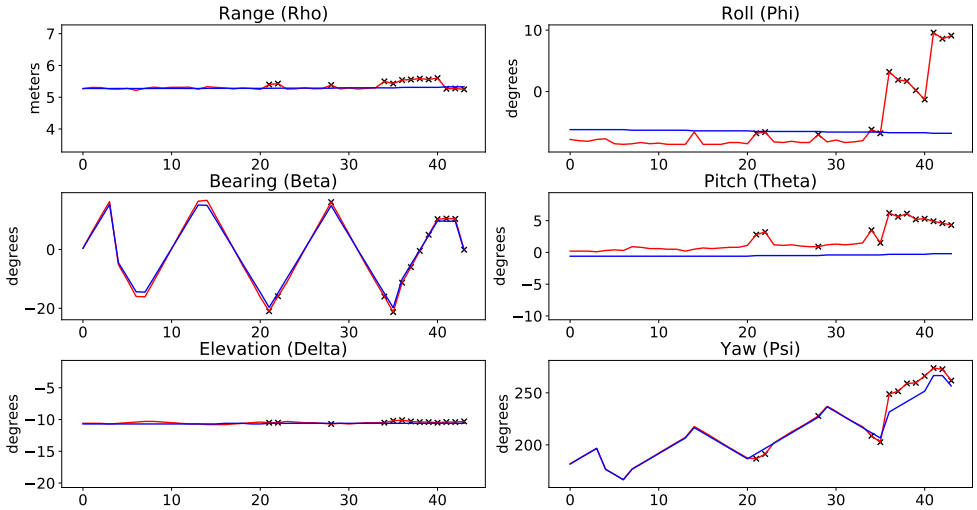


Figure D.14: *5m range, 5m depth, 0.9m elevation offset (Test 27).*

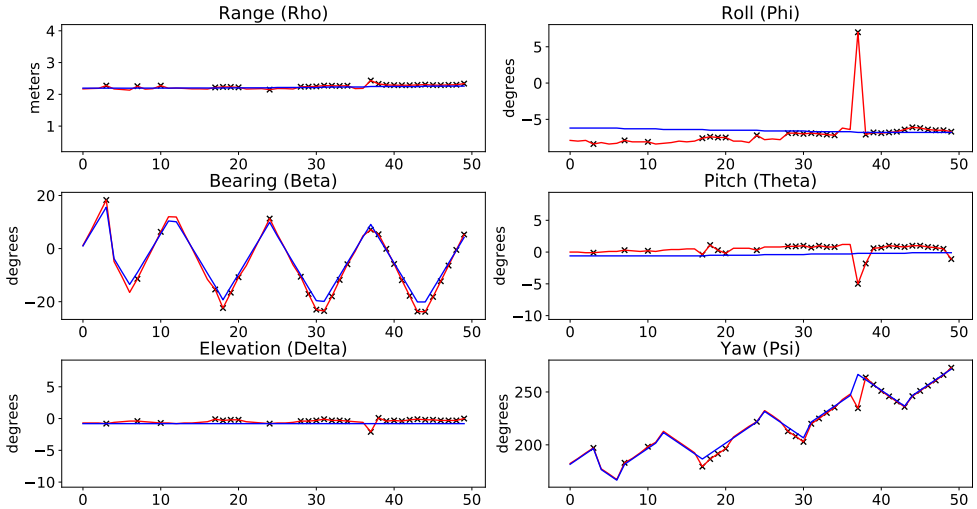


Figure D.15: *2m range, 5m depth, no elevation offset (Test 28).*

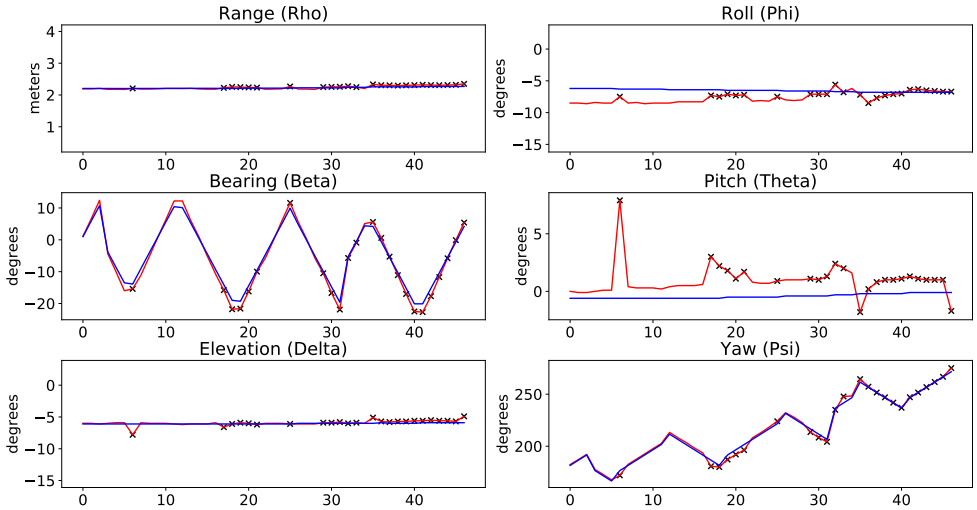


Figure D.16: *2m range, 5m depth, 0.2m elevation offset (Test 29).*

DOCUMENT CONTROL DATA

*Security markings for the title, authors, abstract and keywords must be entered when the document is sensitive

1. ORIGINATOR (Name and address of the organization preparing the document. A DRDC Centre sponsoring a contractor's report, or a tasking agency, is entered in Section 8.) DRDC – Suffield Research Centre Box 4000, Station Main, Medicine Hat AB T1A 8K6, Canada	2a. SECURITY MARKING (Overall security marking of the document, including supplemental markings if applicable.) CAN UNCLASSIFIED 2b. CONTROLLED GOODS NON-CONTROLLED GOODS DMC A

3. TITLE (The document title and sub-title as indicated on the title page.)

Development and Testing of an Underwater Optical Imaging System to Enable Unmanned Underwater Vehicle/Submarine Rendezvous

4. AUTHORS (Last name, followed by initials – ranks, titles, etc. not to be used. Use semi-colon as delimiter)

Giesbrecht, J.; Watt, G. D.

5. DATE OF PUBLICATION (Month and year of publication of document.)	6a. NO. OF PAGES (Total pages, including Annexes, excluding DCD, covering and verso pages.)	6b. NO. OF REFS (Total cited in document.)
September 2018	98	33

7. DOCUMENT CATEGORY (e.g., Scientific Report, Contract Report, Scientific Letter)

Scientific Report

8. SPONSORING CENTRE (The name and address of the department project or laboratory sponsoring the research and development.)

DRDC – Suffield Research Centre
Box 4000, Station Main, Medicine Hat AB T1A 8K6, Canada

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable contract number under which the document was written.)
10a. DRDC DOCUMENT NUMBER DRDC-RDDC-2018-R155	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11a. FUTURE DISTRIBUTION WITHIN CANADA (Approval for further dissemination of the document. Security classification must also be considered.)

Public Release

11b. FUTURE DISTRIBUTION OUTSIDE CANADA (Approval for further dissemination of the document. Security classification must also be considered.)

12. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Use semi-colon as a delimiter.)

Unmanned Underwater Vehicle; submarine; robotic docking; computer vision

13. ABSTRACT/RÉSUMÉ (When available in the document, the French version of the abstract must be included here.)

This document describes the design and test of an optical position sensing system to enable autonomous rendezvous between a submarine and an Unmanned Underwater Vehicle (UUV). The vision system uses LED lighting on the UUV and a monocular camera system to predict the UUV position and pose relative to the camera. Important aspects include the use of a variety of computer vision algorithms changed adaptively over the course of the UUV approach, as well as the use of active control of the camera exposure to enable computer vision. Tests were conducted in harbour water from a barge facility at DRDC Atlantic in Bedford Basin, Nova Scotia using a mechanical ground truth system to verify performance. The system was found to operate in real-time at ranges to at least 13 m over a wide variety of relative positions, poses, and lighting conditions. Typical accuracy was on the order of less than 3% of range, and better than 5 degrees of accuracy in determination of UUV roll, pitch and yaw.

Le présent document vise à décrire la conception et l'essai d'un système optique de détection de la position pour permettre l'amarrage automatique d'un sous-marin à un véhicule sous-marin sans équipage (VSSE). Grâce à un éclairage à DEL à bord du VSSE et un système de caméra monoculaire, le système de vision permet de prévoir la position et l'orientation du VSSE par rapport à la caméra. Parmi les aspects importants du système, mentionnons l'utilisation de divers algorithmes de vision artificielle modifiés de manière adaptative à l'approche du VSSE, ainsi que le recours à un contrôle actif de l'exposition de la caméra visant à permettre une vision par ordinateur. RDDC Atlantique a procédé à des essais de rendement sur une barge, au moyen d'un système mécanique de réalité de terrain, dans les eaux portuaires du bassin de Bedford en Nouvelle Écosse. On a constaté que le système fonctionnait en temps réel jusqu'à une distance d'au moins 13 m, et ce, dans un large éventail de conditions d'éclairage, d'orientations et de positions relatives. La précision type s'est chiffrée à moins de 3 % de la distance et à plus de 5 degrés pour la mesure du roulis, du tangage et du lacet du VSSE.