High-Resolution Underwater Robotic Vision-Based Mapping and Three-Dimensional Reconstruction for Archaeology

Matthew Johnson-Roberson
University of Michigan
Mitch Bryson, Ariell Friedman, and Oscar Pizarro
The University of Sydney
Giancarlo Troni
Pontificia Universidad Católica de Chile
Paul Ozog
University of Michigan
Jon C. Henderson
University of Nottingham

Documenting underwater archaeological sites is an extremely challenging problem. Sites covering large areas are particularly daunting for traditional techniques. In this paper, we present a novel approach to this problem using both an autonomous underwater vehicle (AUV) and a diver-controlled stereo imaging platform to document the submerged Bronze Age city at Pavlopetri, Greece. The result is a three-dimensional (3D) reconstruction covering 26,600 m2 at a resolution of 2 mm/pixel, the largest-scale underwater optical 3D map, at such a resolution, in the world to date. We discuss the advances necessary to achieve this result, including i) an approach to color correct large numbers of images at varying altitudes and over varying bottom types; ii) a large-scale bundle adjustment framework that is capable of handling upward of 400,000 stereo images; and iii) a novel approach to the registration and rapid documentation of an underwater excavations area that can quickly produce maps of site change. We present visual and quantitative comparisons to the authors' previous underwater mapping approaches.

1. INTRODUCTION

Underwater archaeology is an evolving field that leverages advanced technology to improve the accuracy, completeness, and speed at which surveys can be performed. In this paper, we present a new approach to map underwater archaeological sites using both an autonomous underwater vehicle (AUV) and a diver-controlled stereo imaging platform. This approach allowed our team to document a large submerged Bronze Age city at Pavlopetri, Greece, at a previously unachievable scale and resolution.

Historically, underwater archaeological surveys have been conducted by self-contained underwater breathing apparatus (SCUBA) divers using labor-intensive and time-consuming techniques (Bowens, 2009). More recently, coastal archaeologists have used surface vessel–operated sonar (Ballard, 2007), remotely operated vehicles (ROVs) (Ballard et al., 2002; Royal, 2012), and AUVs (Bingham et al., 2010) to cover large areas of the underwater environment in a time-efficient manner. When archaeological sites are located in shallow water (depths as little as 1 m), which frequently occurs in the study of ancient coastal settlements, these platforms are less effective. Large surface vessels are often prevented from operating because of their draft, and larger ROVs and AUVs typically have minimum depth requirements, for example multihull vehicles that offer stability but cannot operate in depths less than 4-5 m (Williams et al., 2012). The interpretation of sonar data employed by these platforms is complicated because of artifacts associated with the geometry and radiometry of acoustic sensing (Mitchell & Somers, 1989; Capus et al., 2008), which is particularly prevalent in shallow waters.

When surveying in shallow waters or at close range to the seafloor, optical sensors are a preferable alternative to sonar as they can operate at higher resolutions but are limited in range (due to the attention of light), therefore requiring many images to be captured to cover a single site. Most optical underwater surveying approaches use mosaicing methods (Foley et al., 2009; Ludvigsen, Sortland,

Direct correspondence to: Matthew Johnson-Roberson, e-mail: mattjr@umich.edu

Journal of Field Robotics 00(0), 1–19 (2016)
View this article online at wileyonlinelibrary.com • DOI: 10.1002/rob.21658

Johnsen, & Singh, 2007; Rzhanov, Cutter, & Huff, 2001), to combine many images into a single spatially contiguous two-dimensional (2D) representation of the environment. Mosaics are useful for visualizing data at scales larger than a single image, but most existing approaches in the literature ignore the 3D structure of the scene (for example, (Pizarro & Singh, 2003)), resulting in geometric distortion and inaccuracies in the mosaic.

In this paper, we develop new techniques for building accurate, broad-scale (>20,000 m^2), and high-resolution ($\leq$2 mm/pixel) maps of shallow (0.5-10 m) underwater environments using optical sensors carried by multiple robotic platforms. We undertook the mapping of a shallow-water archaeological site using two platforms: a stereo diver-rig and a small AUV. These two platforms complemented each other and allowed the researchers to achieve both high-resolution and broad coverage. The diver-rig allowed for the targeted exploration and mapping of high-relief areas that the single-hull, torpedo-shaped AUV was incapable of navigating. In complement, the AUV quickly gathered a high-resolution map of the entire site, providing archaeological context and broad-scale perspective. We developed new postprocessing methods, including image color correction, large-scale bundle adjustment and multi-temporal map registration that allowed for the reconstruction of a large, shallow water Bronze Age city at Pavlopetri, Greece, using sensor data collected by these platforms. Several novel contributions beyond the state of the art in previous approaches in the literature are presented:

- An approach to color correct large numbers of images at varying altitudes and over varying bottom types to produce high-quality images that appear similar to ones taken in air.
- A multiplatform approach to gathering underwater archaeological maps of submerged sites.
- A large-scale bundle adjustment framework that is capable of handling hundreds of thousands of stereo images and producing high-quality, self-consistent poses using both navigation sensors and image constraints.
- A novel method for the rapid documenting of underwater excavations, decreasing in-water time by archaeologists and improving the quality of the documentation of each layer exposed during the excavation process.
- A visual and quantitative comparison of the proposed bundle adjustment approach to the authors' previous simultaneous localization and mapping (SLAM) optimization process (Mahon et al., 2011).
- The reconstruction of approximately 400,000 images into a large-scale model of the submerged Bronze Age city of Pavlopetri.

The paper is organized as follows. The remainder of this section discusses existing mapping approaches, prior art in the field, and the site and its significance. Section 2 discusses the experimental hardware. Section 3 discusses the proposed approach, including the novel image correction method and bundle adjustment. Section 4 presents the results of the excavation mapping, the two field seasons, and finally the comparison to previous mapping techniques. Section 5 discusses conclusions from this work.

1.1. Field Site

Pavlopetri is a shallow submerged prehistoric Bronze Age town lying in 1 to 4 m of water at the west end of the Bay of Vatika in southeastern Laconia, Greece (Harding, Cadogan, & Howell, 1969; Henderson, Gallou, Flemming, & Spondylis, 2011). The remains cover an area of approximately 50,000 m^2 and comprise a network of stone walls, building complexes, courtyards, streets, graves, and rock-cut tombs. The walls are made of uncut aeolianite, sandstone, and limestone blocks and were built without mortar. While some walls remain up to three stones in height, the majority of the site consists of walls and other structures that are only a single stone in height or are flush with the seabed. The dating of the architectural features and surface finds such as ceramics suggest the site was inhabited from at least the Early Bronze Age ca. 3000 BC through to the end of the Late Bronze Age ca. 1100 BC. At its peak, the settlement was likely to have had a population of 500 to 2,000 people.

1.2. Existing Mapping Approaches

Sonar technology is currently the default choice of archaeologists and oceanographers when attempting to map larger areas of the seabed beyond the scale of a single wreck or a single building (Green, 1990; Ballard, 2007; Bowens, 2009). Although the long ranges of acoustic signals allow large areas to be mapped quickly, postprocessing this data poses several challenges. First, the human operator may find it difficult to interpret the data due to relatively low resolution of the sensor (typically grid sizes greater than 10 cm (Capus et al., 2008)). Second, acoustic reflectivity and backscattering create geometric inaccuracies (Mitchell & Somers, 1989; Sakellariou, 2007; Capus et al., 2008). In contrast, while optical sensors can provide high-resolution images in a modality that is easy to interpret, their range is limited to a few meters due to the absorption and scattering of light in water. As a result, short-range optical sensors have to be physically transported close to the underwater features being imaged. This is usually done by attaching the sensors to an remotely operated vehicle (ROV) or similar systems that descend to the seafloor from a boat or similar working platform (Ballard et al., 2002; Bingham et al., 2010; Royal, 2012). Such an approach can result in prohibitive running costs for archaeological budgets and the large size of many ROVs and their support ship make them poorly suited for operations in shallower water. In contrast, small ROVs that are capable of operating in these depths typically lack the precision navigation instruments required to perform structured surveys.

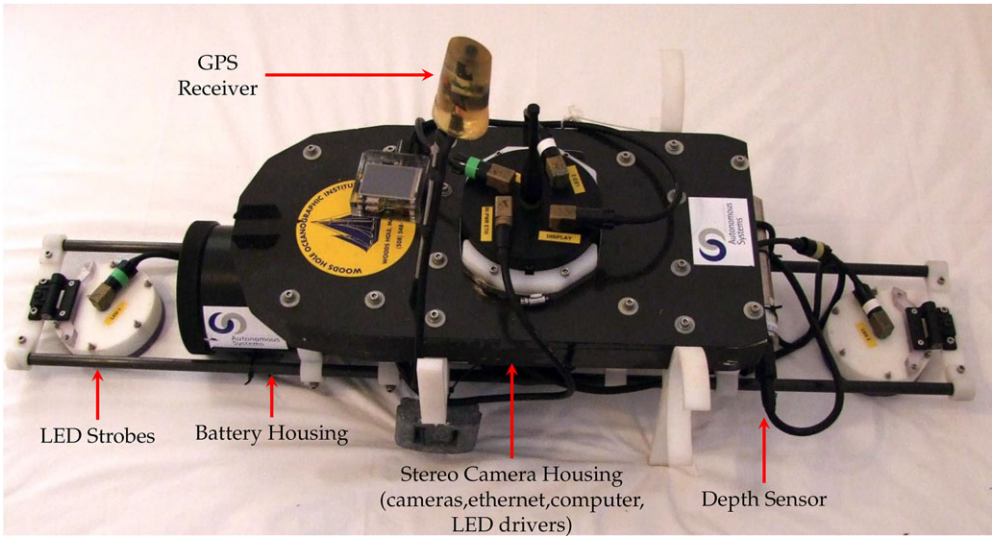

Figure 1. The stereo-vision diver-rig and list of components.

At present, detailed underwater archaeological surveys of sites are typically conducted by SCUBA divers using baselines, fixed site grids, measuring tapes, drawing frames, and photogrammetry (Bowens, 2009). While such approaches can be very effective, they tend to be time consuming, extremely labor intensive, and hampered by adverse weather. In addition, they also scale poorly with the size of the site and the roughness of the underlying terrain. Trench excavation is a common practice in underwater archaeological fieldwork (Henderson, Pizarro, Johnson-Roberson, & Mahon, 2013) involving the successive removal of seafloor sediment in the documentation of buried artifacts, allowing archaeologists to step backward through layers of sediment looking at what was present at different times. The requirement for performing site surveys at each stage of the excavation exacerbates the time taken as each excavated layer must be sketched, planned, and documented fully before continuing to dig.

The recording of submerged sites though the creation of 2D photo-mosaics constructed from overlapping images is common practice (Ballard et al., 2002; Foley et al., 2009). However, the problems with geometrical accuracy over larger areas are widely known. Optical distortion through the water, camera tilt, and variations in the topography of the area being photographed mean that the assumptions for planar mosaics are not generally valid in underwater surveys. As a result, only small groups of photos can be effectively grouped together with mosaics of larger areas, which are less geometrically reliable as errors are compounded the further one builds the mosaic from the center of the first image (Green, 1990).

Attempts to geometrically rectify images prior to assembling mosaics using postprocessing software have had some success (Martin & Martin, 2002), but the primary acquisition of the images still requires the laborious setting up of accurate grids, positioning of reference targets, and the use of rigid bipod frames or towers to ensure pictures are taken from a constant height and the film plane remains horizontal. Equally such techniques remain difficult to use in undulating terrain such as shipwrecks or harbors. In addition, while the proposed techniques all focus on the aqueous environment, there is work showing they can be complemented with an aerial approach to map the nearshore (Bryson, Johnson-Roberson, Murphy, & Bongiorno, 2013a).

2. PLATFORMS

2.1. Stereo Diver-Rig

The Australian Centre for Field Robotics (ACFR) developed a platform to be used in shallow water to collect digital image and sensor data. Referred to as the *diver-rig*, the unit consists of the cameras, lighting, sensors, instrumentation, and power source needed to take high- resolution images of the seabed arranged inside a rigid but highly portable carbon fiber and balsa wood frame (Fig. 1). Specifications for the diver-rig and its sensors are listed in Table I.

The core of the system consists of two highly sensitive digital cameras set up as a stereo pair pointing downward. The stereo pair consists of one color and one grayscale AVT Prosilica GC1380 cameras. They have very sensitive 1.4 M Pixel 2/3″ charge coupled devices (CCDs). The cameras have 8 mm lenses that provide a field of view of approximately 42 degrees × 34 degrees in water. At 2 m altitude, this results in a footprint of approximately 1.5 m × 1.2 m

Table I. Summary of the diver-rig components.

Imaging	
Camera	Prosilica GC1350 12bit 1360×1024 CCD stereo pair
Field of View	42 deg × 34 deg
Stereo Baseline	7.5 cm between cameras
Separation	symmetrical 0.5 m between camera and lights
LED Driver	Gardasoft PP-520F LED driver
Processing	ADL945PC Core 2 Duo PC/104 Plus
Storage	640 GB (8 hr typical @ 4 Hz)
Navigation	
Attitude	Microstrain 3DM-GX1 solid state AHRS
GPS receiver	SPK-GPS-GS405
Depth	Seabird pressure sensor
Communications	
Ethernet	100 BaseT deck cable
Serial	Digikey Portserver (4 serial ports)
External Components	
Lighting	Two 12000 Lumen LED arrays, 4 ms on time
ACFR antenna mast	GPS
Platform specs	
Basic platform	Custom Design
Materials	Carbon fibre and balsa wood composite
Size	60 cm × 80 cm × 20 cm
Mass	12 kg (in air) neutral in water
Propulsion	Human
Batteries	Oceanserver 190 Wh Li-ion pack (3 hr typical)
Maximum Speed	1.0 m/s (diver dependent)
Typical operating speed	0.3-0.4 m/s
Typical Altitude 2 m	
Endurance	3-4 hr
Depth rating	150 m

and a spatial resolution of ∼ 1mm/px. The two cameras are triggered simultaneously by a microcontroller typically at 2Hz, providing three to five views of the same scene point at typical speeds and altitudes.

Two light emitting diode (LED) strobe units were used, 0.5 m fore and aft of the down-looking stereo cameras. The cameras and strobes were used in an auto-exposure mode that would adjust the exposure time to achieve an average intensity of 30% of the range value. This allowed operating under a wide range of lighting conditions while maintaining similar illumination levels. The secondary instruments consisted of a pressure sensor (depth), surface global positioning system (GPS) receiver, and a solid-state inertial measurement unit (IMU) (for attitude). The design of the system was based on the same component array used and tested in autonomous underwater vehicle (AUV) and ROV configurations by the ACFR on previous oceanographic research missions (Bryson, Johnson-Roberson, Pizarro, & Williams, 2013b; Johnson-Roberson, Pizarro, Williams, & Mahon, 2010). The diver-rig differs in that it uses a subset of secondary instruments and thus is more heavily dependent on visual odometry and GPS when compared with the AUV. From a locomotion standpoint, the form factor of the platform was designed to be manipulated by divers as opposed to being attached to an AUV or ROV.

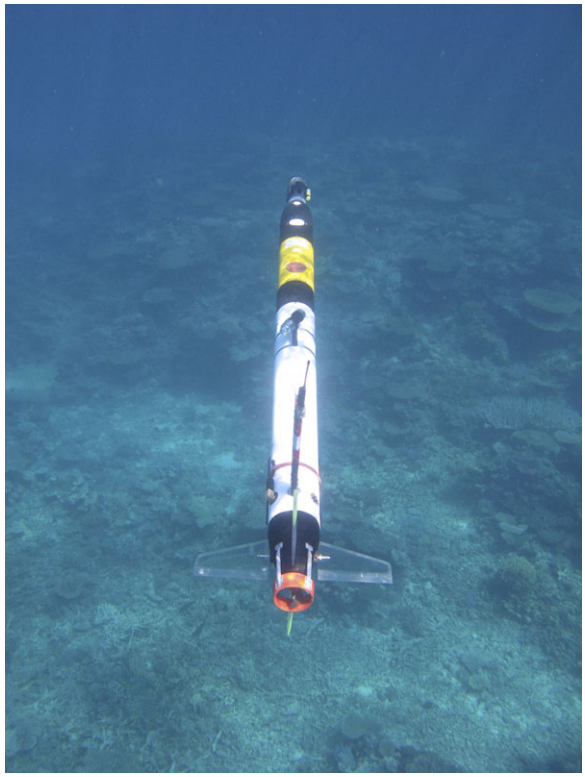

Figure 2. Iver2 AUV used in Pavlopetri mapping.

2.1.1. Survey Class AUV with Stereo Imaging Package

In complement to the diver-rig system, a torpedo-shaped AUV was also deployed for the field experiments. The AUV is a modified version of the OceanServer Technology Iver2 (see Fig. 2), carrying a stereo imaging section similar to the one in the diver- rig. Specifications for the AUV and its sensors are listed in Table II.

The use of an AUV greatly increases the coverage rate of the surveyed area. Mapping the ancient city of Pavlopetri took less than 24 hr of in-water time to complete. At Pavlopetri, the entire site was at a shallow enough depth, allowing for the use of GPS while the seafloor was still in imaging range.

Table II. Summary of the Iver2 components.

Imaging	
Camera	Prosilica GC1350 12bit 1360×1024 CCD stereo pair
Field of View	42 deg × 34 deg (down-looking)
Stereo Baseline	9.0 cm between cameras
Separation	0.5 m fore and 1.2 m aft between camera and lights
LED Driver	ACFR LED driver
Processing	ADLGS45 Core 2 Duo PC/104 Plus
Storage	640 GB (8 hr typical @ 4 Hz)
Navigation	
DVL	RDI Navigator ADCP 600kHz
GPS receiver	SPK-GPS-GS405
Attitude	Oceanserver Compass + pitch and roll
Depth	Oceanserver/YSI pressure
Communications	
RF Modem	900 MHz
Ethernet	100 BaseT deck cable
External Components	
Lighting	Two 12000 Lumen LED arrays, 4 ms on time
ACFR antenna mast	GPS, RC antenna, Iridium tracking
Platform specs	
Basic platform	Ocean-Server Iver2 42-inch
Materials	Carbon fibre, Delrin and Aluminium sections
Size	255 cm × 14.7 cm + (keel, fins,mast)
Mass	46 kg
Propulsion	160 W direct drive brushless DC thruster
Manual RC override	
Batteries	Oceanserver 760 Wh Li-ion pack (8 hr typical)
Maximum Speed	2 m/s
Typical operating speed	1 m/s
Typical Altitude 2 m	
Endurance	4-5 hr @ 1 m/s
Depth rating	100 m

3. POSTPROCESSING AND MAP RECONSTRUCTION

3.1. Image Quality and Correction

Compared to terrestrial environments, light behaves much differently underwater and thus poses several practical issues for capturing high-quality images. In particular, water absorbs light as it passes through–a process called attenuation (Duntley, 1963). Attenuation reduces the intensity of light in relation to the distance traveled. For this reason, sunlight, commonly used as the primary lighting source in terrestrial photogrammetry, is typically not strong enough to illuminate scenes below depths of approximately 20 m. This necessitates the use of artificial lighting onboard an underwater imaging platform when operating at depth.

The attenuation of light underwater is frequency-frequency dependent; red light is attenuated over much shorter distances than green or blue light, resulting in a change in the observed color of an object at different distances from the camera and light source. In the context of optical mapping, the color and reflectivity of objects is significantly different when imaged from different camera perspectives and distances. This can cause problems for computer-based techniques regarding the matching and alignment of image data based on color intensities. This is because an image patch being tracked on a moving camera violates the brightness constancy constraint (BCC) that underlies most image matching algorithms. To address this, we propose a novel image clustering approach to the problem of underwater image correction. The approach of (Johnson-Roberson et al., 2010) demonstrated the viability of using the gray-world assumption (Buchsbaum, 1980) to correct illumination variations in underwater images.

One major assumption of that work was that the images were taken at relatively constant altitude. With a single platform, in this case the Sirius AUV (Williams et al., 2012), that assumption was reasonable as the Doppler velocity log (DVL) allowed for accurate altitude control over mild seafloor relief. Here in this application that assumption was violated. The combination of diver-gathered imagery and a very shallow deployment location led to much larger variance in altitude across the data set. As the goal of the fieldwork was to map a large extent of the site, the image correction across all the images needed to be consistent. Another challenge presented by the large-area coverage of this site was the diversity of bottom type. Some areas were dominated by seagrass, others by sand, and still others by large stones. This drastically changed the distribution of colors in the images and led to suboptimal results. The authors and others have proposed more complex attenuation correction processes (Bryson, Johnson-Roberson, Pizarro, & Williams, 2012; Sedlazeck, Koser, & Koch, 2009). However, the sheer scale of this data necessitated the development of a novel approach focused on scalability and efficiency.

The gray-world correction approach assumes a single unimodal Gaussian distribution of intensities. In our prior approach in (Johnson-Roberson et al., 2010), this assumption was applied to all images across a deployment. As reconstructing larger areas became possible, the varying bottom types more strongly violated the underlying assumptions of the gray-world approach. Large missions display a multimodal distribution of intensities and colors. To attempt to prevent the flattening of these mixed color pallets in a single correction, we propose a two-layered clustering approach. This approach splits images, first into bands of discrete altitude ranges (we selected 10 cm divisions from 1.0-4.0 m, which still left thousands of images in each band) and then into clusters based on bottom type.

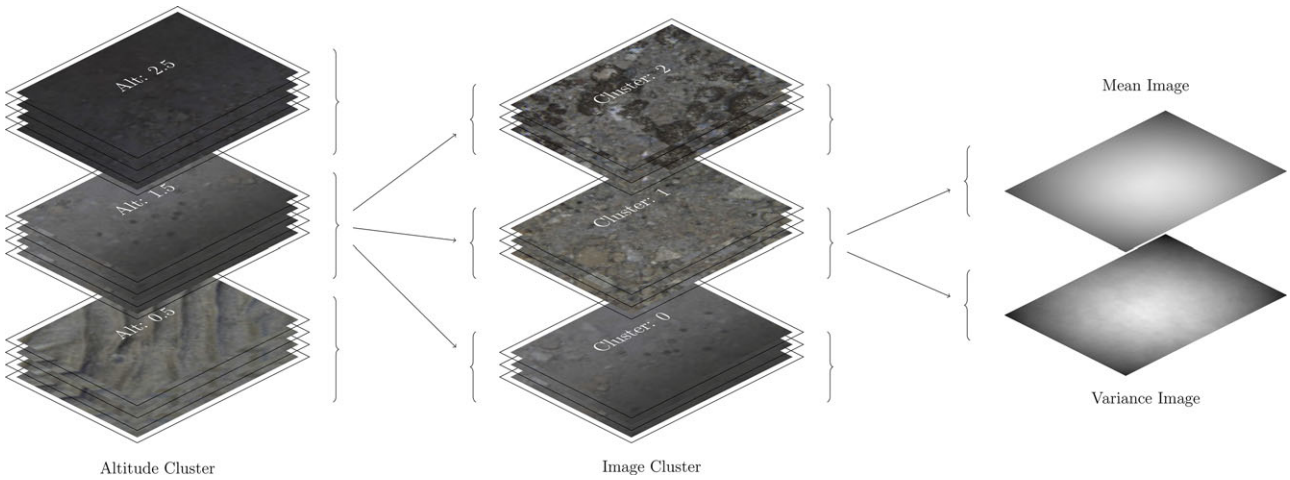

Figure 3. A diagram depicting the two-level clustering process used in image correct, first a division by imaging altitude and then a division by estimated bottom type using a VDP clustering algorithm (Steinberg et al., 2011, 2013). After the creation of all these subsets, the mean and variance for each pixel location in each subset is computed. A gain and offset is then calculated using the standard gray-world algorithm (Buchsbaum, 1980).

This is done using a VDP clustering algorithm (a diagram appears in Fig. 3) (Steinberg, Friedman, Pizarro, & Williams, 2011; Steinberg, Pizarro, & Williams, 2013).

The VDP clustering is performed on the grayscale images using scale-invariant feature transform (SIFT) features, which capture more of the texture of the bottom than the illumination characteristics. The VDP clustering is unsupervised and selects a number of clusters based on the data. The result is there are a variable number of levels in the second division of each altitude band (from three to eight in this data set). For each subset, a mean and a variance image is calculated. This two-leveled approach increases the homogeneity of images from which a pixel correction is calculated. A gain and offset for each pixel position and channel is then calculated to transform the distribution to be the same across each of these subsets. The results of the uncorrected and corrected images appear in Fig. 4. Another important note is this process occurs on the cameras' 12-bit images, increasing the dynamic range in the correction process. While this cannot completely ensure that no information is lost to clipping, the large dynamic range goes a great way to ameliorating this problem, particularly on light sediment with dark artifacts or vice versa.

3.2. Bundle Adjusted Map Reconstruction

To obtain the topographic relief and overlaid photo-textured reconstruction of the site, all imagery and sensor data collected were processed using SLAM and mesh reconstruction algorithms. Real-time navigation using the available sensor data was then used to seed an offline bundle adjustment (Triggs, McLauchlan, Hartley, & Fitzgibbon, 2000) pipeline. Fig. 5 depicts an architectural block diagram of our implemented formulation with each step discussed in more detail in the following sections.

3.3. Real-Time Navigation

A real-time navigation system using an extended Kalman filter (EKF) was used to combine sensor data from GPS (while on the surface), the DVL, depth, and attitude sensor data into initial estimates of the 6 degrees of freedom (DOF) pose of the platform corresponding to times at which images were captured by the stereo camera system. This approach benefits from the fusion of multiple sensors, each with uncertainty models that allow for their probabilistic integration (Thrun, Burgard, & Fox, 2005). The combination of these sensors provided a position accuracy of approximately +/- 2 m horizontal and 10 cm vertically and an orientation accuracy of approximately $+/-2^{\circ}$, which was insufficient for use in spatially registering stereo image pairs directly, but sufficient for path following that produced mostly complete coverage and also sufficient in providing the initial seed parameters to the offline bundle adjustment procedure described below.

3.4. Factor-graph Formulation and Stereo Matching Search

Our formulation for recovering an optimized pose and scene structure is inspired by the factor-graph representation of the SLAM problem (Dellaert & Kaess, 2006). This method leverages the inherent sparsity of a least squares problem corresponding to a maximum a *posteriori* (MAP) estimate of the vehicle trajectory and 3D structure. In the factor-graph formulation, the unknowns correspond to

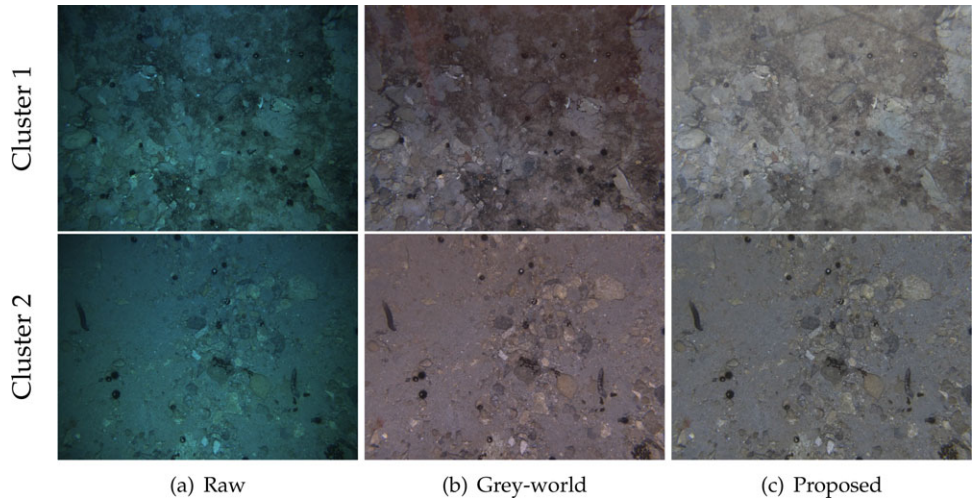

Figure 4. An example of the proposed lighting correction. Each row displays the processing of a single image from a unique cluster produced using a VDP. Note both unique images came from the same altitude band. The original images appear in column (a). Note the strong vignetting (dark corners), and a blue-green hue caused by the rapid attenuation of red light. The standard gray-world correction is shown in column (b). Note the reddish tinge. This is the result of integrating of many images with varying bottom-type reflectivity. The correction is too aggressive in increasing the red component for lighter images. Finally, the result of the color balancing process described in Section 3.1 is shown in column (c). Note the overall consistency of both images and the removal of most vignetting.

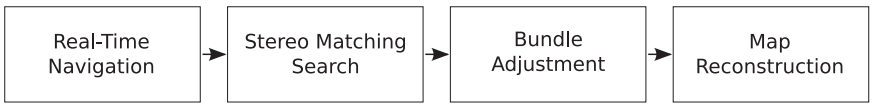

Figure 5. Overview of the implemented pipeline for vision-based mapping and 3D reconstruction.

variable nodes, while the measurements are represented by factor nodes.

For this work, the variable nodes consist of

(i) 6-degree of freedom (DOF) vehicle poses,
(ii) a 6-DOF rigid transformation from the vehicle frame to the camera frame, and
(iii) 3D points that generate a matched speeded-up robust features (SURF) (Bay, Ess, Tuytelaars, & van Gool, 2008).

The nodes positions were initialized to the position reported by the real-time vehicle navigation filter using GPS, DVL and attitude data.

The factor nodes consist of

(i) full 6-DOF priors corresponding to initial position and extrinsic camera calibration,
(ii) depth, pitch, and roll (ZPR) priors,
(iii) odometry constraints between successive vehicle poses (if available), and
(iv) stereo reprojection error.

A diagram of this formulation is shown in Fig. 6.

Once we have encoded 6-DOF vehicle poses information on a factor-graph format, we calculate from the imagery the matching 3D SURF features positions ($\mathbf{l}_i$) and the corresponding factor nodes matching keyframes to be encoded in the factor-graph used in the optimization stage. We use prior Euclidean distance between each node to conservatively limit the stereo matching search to only adjacent nodes (< 2.5m as the GPS and DVL provided good starting odometry).

We compute the relative motion constraint based on the following pipeline. We acquired and time stamped the images at 4Hz. First, SURF features and descriptors (Bay et al., 2008) are detected in grayscale calibrated images using the OpenCV library (Bradski & Kaehler, 2008). Then the features from both stereo images are matched by appearance, yielding a putative set of corresponding 3D points.

To avoid including incorrect data associations (outliers) in the optimization stage, we calculated the relative motion from matching 3D extracted features and keep the ones more consistent with the model. Similar to common stereo-based visual odometry techniques (Maimone, Cheng, & Matthies, 2007), we used a 3D random sample consensus (RANSAC) algorithm to reject outliers. Using Arun's

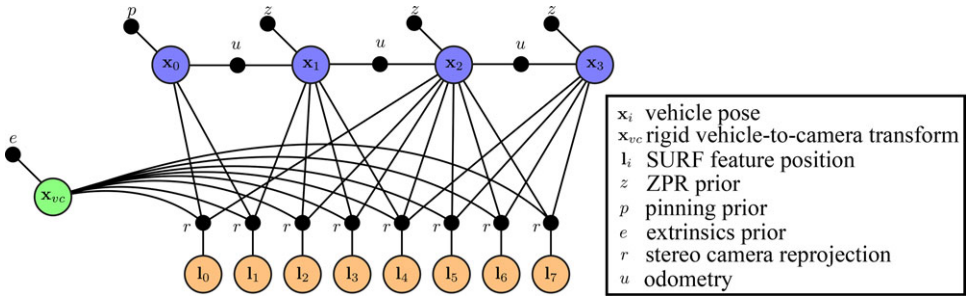

Figure 6. Small example using a factor-graph to graphically represent the estimation of pose (blue nodes) and 3D structure (orange nodes). Besides the usual ZPR and odometry measurements, we use ternary factors connected to 3D features, $\mathbf{l}_0 \ldots \mathbf{l}_7$, to also constrain the estimate of the vehicle-to-camera rigid body transformation, $\mathbf{x}_{vc}$. The final estimate of the vehicle poses, $\mathbf{x}_0 \ldots \mathbf{x}_3$, are fed into a mesh reconstruction and texturing algorithm.

singular value decomposition (SVD) based approach, reported in (Arun, Huang, & Blostein, 1987; Umeyama, 1991), three randomly selected 3D feature matches are selected at a time to fit a relative transformation between the two poses. To identify inliers, we use a simple distance threshold of 1 cm. The transformation that yields the most inliers was refined and used to identify the final set of outliers.

Finally, a factor representing reprojection error is added between all inlier feature nodes and the vehicle pose nodes from which they were observed, as shown in Fig. 6.

3.5. Bundle Adjustment

A bundle adjustment procedure was employed to combine image feature matches to produce an accurate and globally consistent set of pose estimates, which are required for the remainder of the 3D reconstruction pipeline. The procedure optimized pose estimates by minimizing the stereo reprojection errors associated with the factor-graph representation. Each factor described in Section 3.1 was implemented as a cost function term, with the addition of an M-estimator for the stereo reprojection errors. In particular, we used the Huber loss function (Huber, 1964), denoted as ρ with a scaling parameter equal to unity. The resulting MAP estimate was solved using the popular Ceres Solver (Agarwal et al., 2014). Formally, we define the cost function as:

$$\begin{aligned}&\Psi(\mathbf{x}_i, \mathbf{x}_{vc}, \mathbf{l}_j; \mathrm{K}^L, \mathrm{K}^R, \mathbf{x}_{LR}) \\ &\quad = \rho\Big(\|\underline{\hat{\mathbf{u}}}_{ij}^L - \mathbf{u}_{ij}^L\|_{\Sigma_l}^2 + \|\underline{\hat{\mathbf{u}}}_{ij}^R - \mathbf{u}_{ij}^R\|_{\Sigma_l}^2\Big), \end{aligned} \tag{1}$$

where

$$\begin{aligned}\hat{\mathbf{u}}_{ij}^L &= \mathrm{K}^L\left(\mathrm{R}_i^L\mathbf{l}_j + \mathbf{t}_i^L\right) \\ \hat{\mathbf{u}}_{ij}^R &= \mathrm{K}^R\left(\mathrm{R}_i^R\mathbf{l}_j + \mathbf{t}_i^R\right),\end{aligned}$$

and $\underline{\mathbf{u}}$ denotes the dehomogenization of a vector $\mathbf{u}$. K^L and K^R are the camera calibration matrices for the left and right cameras in the stereo rig, respectively. R_i^L and $\mathbf{t}_i^L$ are the rotation and translation corresponding to the pose of stereo rig's left camera at time i. Similarly, R_i^R and $\mathbf{t}_i^R$ correspond to the right camera pose at time i, which is taken by compounding the left stereo pose with the transformation from the left camera to the right camera, $\mathbf{x}_{LR}$, which we assume known from the stereo camera's calibration. Σ_l is the covariance of the observed pixel location. For our experiments, we take $\Sigma_l = 4\mathrm{I}_{2\times 2}$.

Table III. Bundle adjustment results across approximately 360,000 Iver2 images. The *Parameter blocks* and *Residual blocks* entries denote the number of variable nodes and factor nodes, respectively, used in the factor-graph representation from Section 3.4.

Graph size	
Parameter blocks	5749931
Parameters	17633425
Residual blocks	32316872
Residuals	129115870
Costs	
Initial	1.703516e+10
Final	1.163570e+08
Change	1.691880e+10
Time	
Preprocessor	670s
Residual evaluation	229s
Jacobian evaluation	3918s
Linear solver	42441s
Postprocessor	17s
Total	48958s

The size of the optimization problem, along with performance information, was evaluated on a distributed-memory cluster, consisting of eight consumer-grade desktop computers, each with a four-core 1.8 GHz central processing unit (CPU). The performance statistics for running the bundle adjustment on approximately 360,000 Iver2 gathered images covering 26,600 m^2 at a resolution of 2 mm/pixel is shown in Table III. The *Parameter blocks* and

(a) Without blending (b) With blending

Figure 7. Texture blending example. Red circles (a) identify inconsistencies before blending. Green circles (b) indicate improved blended textures.

Residual blocks entries denote the number of variable nodes and factor nodes, respectively, used in the factor-graph representation from Section 3.4. The parameter blocks are stacked into a single column vector which has length given by the *Parameters* entry. The same holds for the *Residuals* entry.

3.6. Three-Dimensional Map Generation and Visualization

Three-dimensional reconstructions are built using the system described in (Johnson-Roberson et al., 2010; Johnson-Roberson, Bryson, Douillard, Pizarro, & Williams, 2013). Corner features are extracted from each stereo image pai, and triangulated to calculate their positions relative to the cameras. The point clouds are converted into Delaunay triangulated meshes and registered in a global reference frame using the SLAM estimated camera poses. The individual surfaces are fused into a single mesh using volumetric range image processing (Curless & Levoy, 1996). Textures are produced by projecting the images onto the mesh structure, using blending over multiple spatial frequency bands (Burt & Adelson, 1983). Fig. 7 illustrates an example of the texture blending process, in which visually consistent textures are produced in the presence of inconsistent illumination and the small registration errors that arise when projecting images onto an approximate 3D structure.

Because many thousands of images can be acquired during a survey, the quantity of data in the final 3D reconstruction can be larger than the available memory on a computer used for visualization. A paged level of detail scheme (Clark, 1976) is used, in which several discrete simplifications of geometry and texture data are generated. Visualization can then be performed efficiently by reducing the complexity of components in the displayed 3D scene proportionally to their viewing distance or relative screen space.

3.6.1. Multimission Processing and Registration for Change Analysis

Archaeological investigations typically involve excavations that necessitates site mapping and survey to be repeated at the various stages of a dig. Here we present an approach that allowed for the registration of independently bundle-adjusted missions to look at change across time. The accuracy of the GPS measurements used to globally geo-reference each map in the previous sections (approximately $+/-$ 1 m) meant that, when subsequent dives were performed in the same area, the resulting maps were not well aligned. Another cause for such discrepancies can be tidal changes. A postprocessing procedure was developed to co-register maps reconstructed using multiple dives repeated in the same area into a single reference frame using matching sets of SIFT feature keypoints extracted from the texture map layer of each map (Lowe, 2004).

The registration parameters were computed in two stages. In the first stage, SIFT feature points were extracted for each texture map layer to be co-registered. For each and every pair of maps, SIFT feature points were matched based on SIFT descriptors, such that for N repeat maps there were $M = \frac{1}{2}(N^2 - N)$ sets of feature correspondences. The vertical position of each matching feature point in each correspondence set was computed using the topographic height layer of the map. A 6-DOF rigid transformation (consisting of a translation and rotation) that aligned two maps into one reference frame was computed for each of the M pairs using a 3D RANSAC algorithm to compute inliers and outliers of matched 3D feature points. The RANSAC algorithm used three randomly selected feature matches at a time and an iterated least squares procedure to find a sample registration between the map pair. The residual errors associated with the remaining feature matches were recorded and the number of matches that satisfied a maximum error of 5 cm (i.e., the inliers associated with the sample registration) was used to assess the candidate registration. After a sufficient number of trials were run, the registration that resulted in

Figure 8. The total station surveying system. An operator on land records measurements (inset), while divers or snorkelers manage a prism pole. The operator must signal divers in the water to coordinate the mapping (reproduced from (Henderson et al., 2013)).

the highest inlier count was used to record not the registration parameters themselves but only which feature matches from each pair of missions were considered inliers. We created separate missions for the excavation and the proposed process is performed on the entirety of those models. In addition, while a minimum of only three points are required to produce a registration, we set our inlier threshold to at least 100 features to ensure robust matching.

In the second stage of the registration algorithm, all of the inlier matches from each and every pair of maps was used in a global nonlinear least squares procedure that jointly estimated the complete set of registration parameters ($6(N-1)$ states in total) from each and every inlier match in all M pairs of maps.

4. RESULTS

In this section, we will present the results of the two field deployment, spaced a year apart with a total of 3 weeks working in the field, highlighting the power of using a stereo mapping platform to aid in archaeological data collection. We examine the production of site plans through broad area survey and present the resulting maps that span tens of thousands of square meters of the seafloor.

In 2009, the University of Nottingham, the Ephorate of Underwater Antiquities, and the Hellenic Centre for Marine Research (HCMR) started a 5-year collaborative project to study the submerged Bronze Age town. The Pavlopetri Underwater Archaeology Project aimed to identify when the site was occupied, how it functioned as a harbor town, how it came to be submerged, and how maritime trade was articulated there. The project had two main phases that made up a comprehensive underwater survey of the submerged remains to document the site and its current condition (2009–2010), followed by three seasons of targeted underwater excavation (2011-2013).

Archaeologists from the University of Nottingham started surveying the site using a total station equipped with data-logging software in 2009. The shore-based total station system was used to take 3D points from a detail pole equipped with a prism, held by divers at carefully chosen locations on the site (Fig. 8) (Henderson & Burgess, 1996). Using bubble levels on the detail pole, the maximum error using this system was found to be less than 5 cm at a full pole extension of 5 m. While the total station mapping continued in 2010, the Australian Centre for Field Robotics (ACFR) was invited to test and evaluate underwater vision-based mapping methods for archaeological applications.

4.1. Diver Rig and AUV Fieldwork (2010 and 2011)

During the 2010 field season, a total of 47 dives were performed using the diver-rig over a period of 10 days, gathering more than 135,000 pairs of stereo images and surveying approximately 40% of the 30,000 m^2 site. A typical dive surveyed a grid area of approximately 15 by 10 m, and took on the order of 1 hour to complete. The process involved three divers and used guidelines to enable complete coverage.

Figure 9. Vehicle trajectories: the blue line denotes the path of the Iver2, while the red line denotes the path of the diver-rig missions. The vast majority of the site was fully covered using the two technologies by the end of the 2011 field season.

The AUV was first used in 2011. That year engineering trials were performed to test surveying strategies and tune depth and heading control for the AUV, as well as refine some of the mission control, logging, and postprocessing software pipeline in the week leading up to the 2011 field season. During the 2011 field season, over 600,000 images were gathered with the AUV covering the remainder of the known extents of the site. Ultimately, the combination of AUV and diver-rig allowed for the complete coverage of the entire site to be completed in a 3-week field effort. During this time, the diver rig was also used to produce a multitemporal map data set of an excavation trench over a period of 9 days of digging. Each day after the diver-based excavation activities were performed, the diver-rig was used to image and map the trench area so the archaeologist could continue digging the following morning. Snorkels were used during diver-rig deployment owing to the water depths of 1-3 m. To our knowledge, this volume of imagery represents the highest density and extent of optical stereo coverage of any underwater site up until now. A depiction of the aerial imagery of the coastline with the paths of the two vehicles for all dives in each deployment overlaid appears in Fig. 9. Special care was put into designing achievable plans for the AUV, which minimize turning over areas of interest. To this end, so-called Zamboni patterns were employed. These are the patterns employed by ice hockey rink grooming machines, in which no sharp turns are needed to densely cover an area. These patterns work well for vehicles with limited turning radii at the expense of duplicate coverage.

AUV operations were performed at night and consisted of shift-based deployment and supervision from shore. Surveying at night allowed for the mitigation caustic lighting effects produced by refraction through ripples and waves on the ocean surface. In addition to making the images more difficult to interpret, such effects reduced the performance of the feature extraction and matching algorithms and should be avoided. Following sunset, the robot would be placed into the water, its mission programmed, and operations started. Typically, one to two people could take responsibility for monitoring the progress of the robot overnight and the vehicle would be retrieved at dawn, its data downloaded and batteries recharged during the day. The diver-rig operations were performed in the early morning or near dusk, which dealt with issues of direct sunlight while not requiring the more complicated logistics of night diving.

Each day a first pass of data processing was completed in order to assess the success of the previous days deployment. Going from the initial retrieval of the vehicle to a semifinalized 3D reconstruction could be completed in several hours. This rapid turnaround enabled debugging and improvement of the previous day's operations and results. Night operations enabled a clear separation between the archaeologist's work (which involved people and equipment in the water) and robotic data gathering. Traditional archaeological methods could be employed during the day, while experimental tools like the AUV could be tested without fear of interfering with the total station work.

4.1.1. 2010 Diver-Rig Field Season Results

Overhead views of the 3D model for three combined survey boxes are shown in Fig. 10. Walls can be clearly seen in both the depth-colored view in Fig. 10 (a), and the texture mapped view in Fig. 10 (b). In addition, the visible structures agree well with the total station chart of the area shown in Fig. 10 (c).

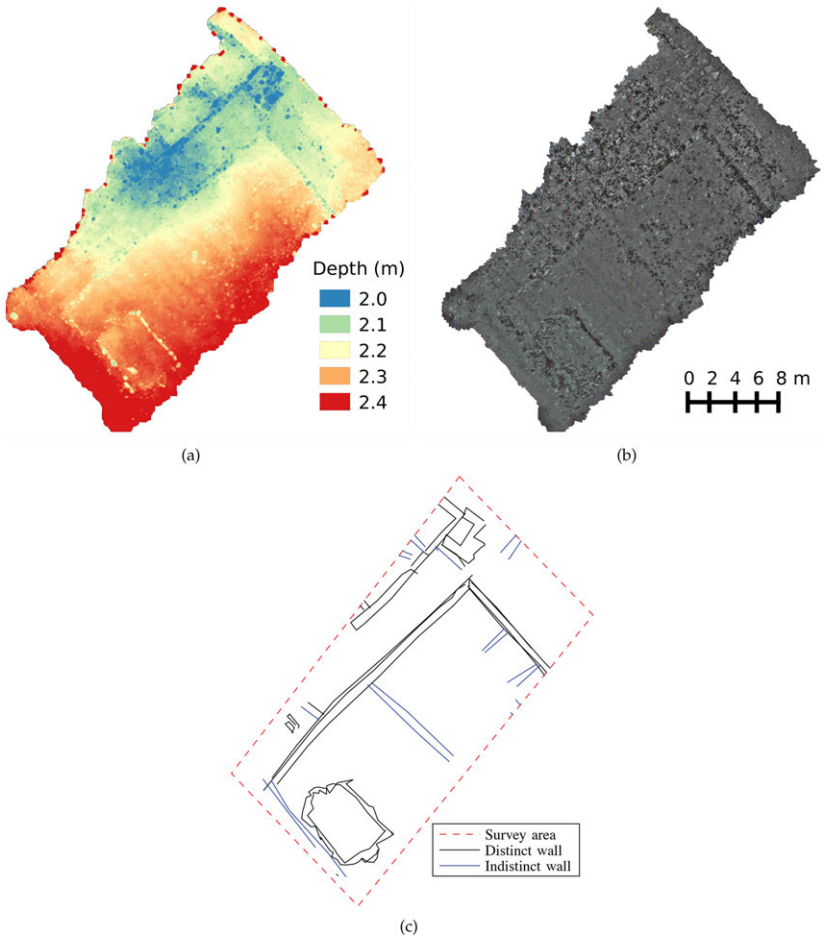

Figure 10. Map of the survey area of from three separate dives using the diver-rig. Remains of walls are clearly visible in both the depth-colored view (a) and the texture mapped view (b). The corresponding map produced by the total station system (c) shows good agreement in the layout of the features of the site.

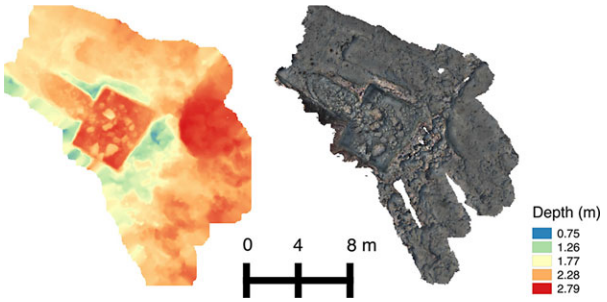

Figure 11. Chamber tomb mesh. The chamber is approximately $4 \times 5\text{m}^2$ in size, and has been cut to a depth of 3 m from the surrounding rock.

Fig. 11 displays overhead views of a mesh created from the survey of a tomb located on the northeast edge of the town. The tomb was cut from the surrounding rock, producing a structure with large depth variations. Some holes in the mesh are present because the survey was performed without the use of any grid markers or navigation aids.

4.1.2. 2011 Diver-Rig Excavation Results

A 9-day series of repeated maps of excavation trench were reconstructed and co-registered (as described in Sec. 3.6.1), are shown in Fig. 12 where both the visual texture and the shaded relief appear, respectively. The resulting registrations had a horizontal residual registration error of approximately 2 cm and a vertical residual registration error of approximately 1 cm. The trenches were approximately 7×4 m, and they could be covered by one person without guides completely in approximately 15 min for each dive.

4.1.3. 2011 AUV Field Season Results

This year's results supplemented the 2010 diver-gathered data with the large-scale AUV imagery. The maps

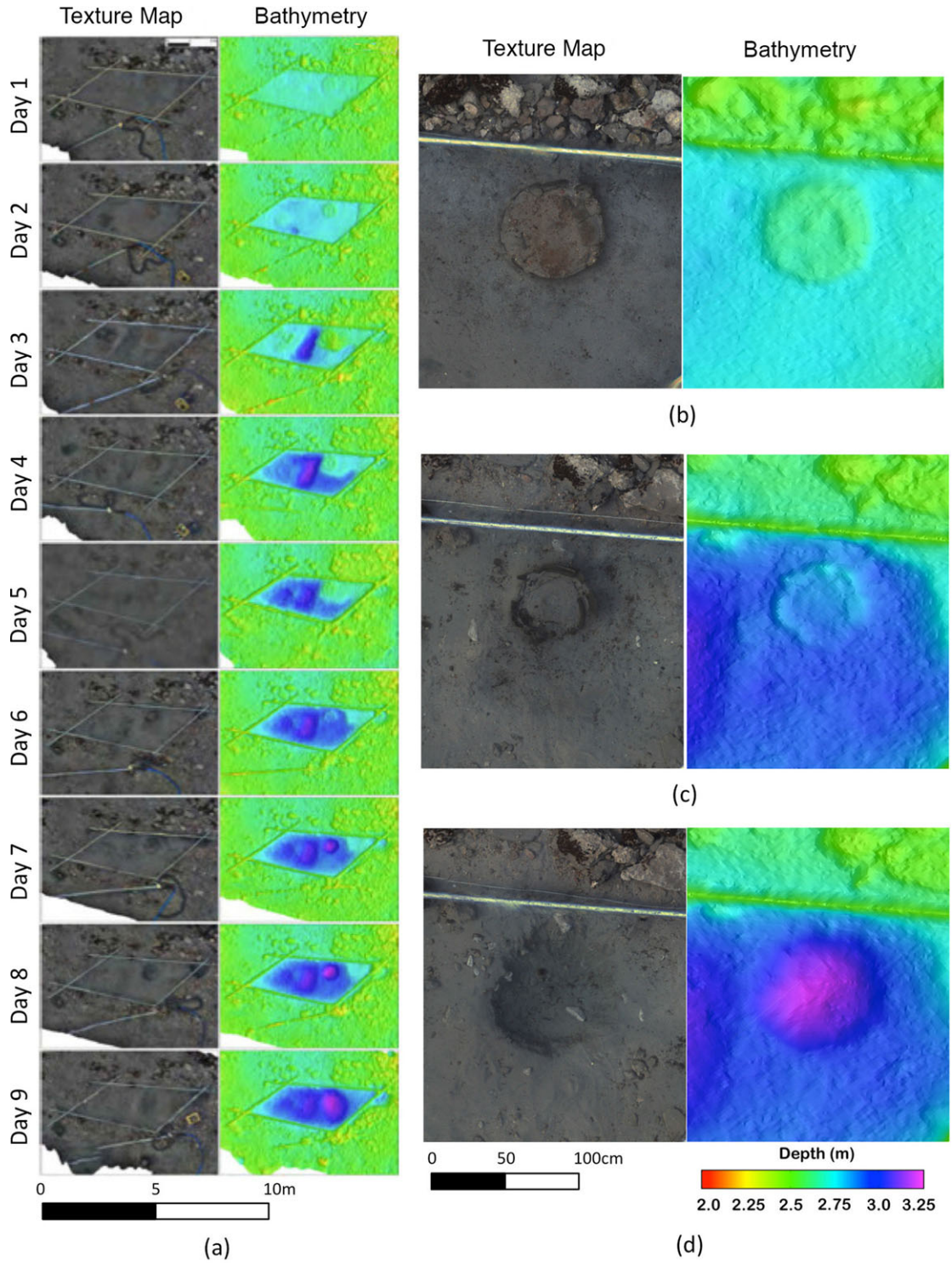

Figure 12. A series of reconstructions of an excavation site. (a) Each row represents a new day of excavation, one layer deeper as they proceed down the page. The left column is the visual image texture, while the right column is the shaded topographic relief. (b, c, d) Zoomed-in imagery and topographic layers illustrating the excavation of a large prehistoric storage jar *(pithos)* on (b) Day 2, (c) Day 6, and (d) Day 8.

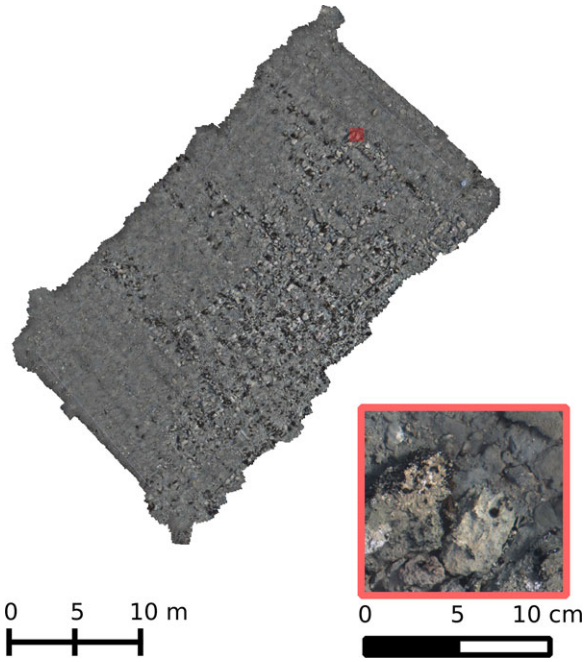

Figure 13. Map from a representative single survey box, which will be used to compare the authors' previous approach to localization with the one proposed in this paper.

generated in the 2011 field season are shown in Fig. 17 with the corresponding depth map shown in Fig. 18. A summary of the size and performance of the proposed method is shown in Table III.

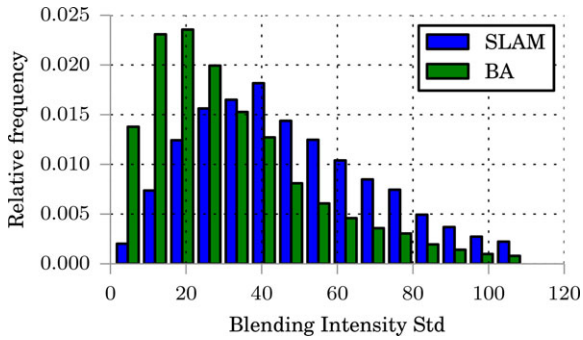

Figure 15. Histogram of intensity deviation for the results shown in Fig. 14. The y-axis displays normalized frequency counts across the whole model. Note the greater quantitative consistency of the the proposed approach (titled BA). There are many more occurrences of great pixel deviation in the authors' previous work (titled SLAM) which does not have global bundle adjustment applied.

4.2. Visual and Quantitative Comparison to Previous Work

To assess the improvements of the bundle-adjusted result over the author's previously published approaches, we present both a visual and quantitative comparison of the previous SLAM work of Mahon et al. (Mahon et al., 2011) to the technique presented here. We selected a representative survey box that appears in Fig. 13 to perform the comparison over. For a qualitative understanding of the results of this work, we show the standard deviation of the pixel values across the box 11 mesh in Fig. 14. This deviation captured

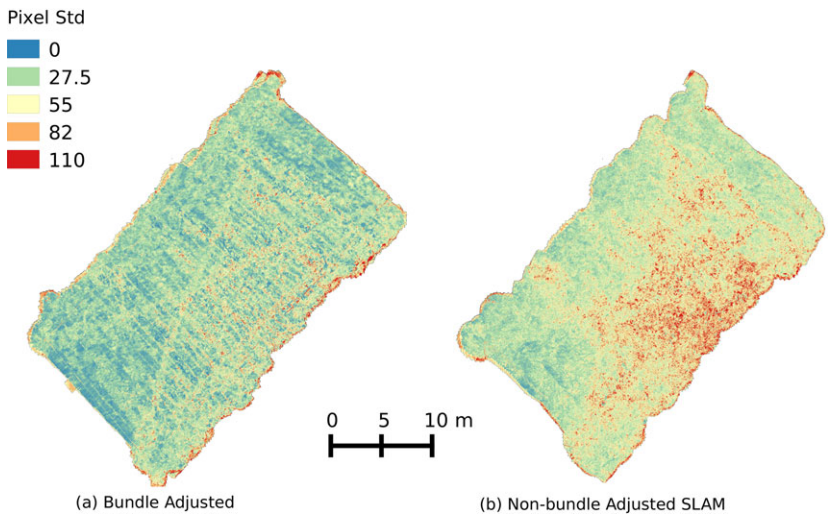

Figure 14. These figures encode the standard deviation of pixel intensities from the reprojection of pixels from independent poses. The result here is on intensity of the image for clarity in the visualization. The individual color channel results display a similar trend. High deviation suggests that the poses are less self-consistent and disagree on the intensity of points in the scene. Blue values denote consistent pixel intensities, while red shows relatively large variation. There is a large region of inconsistency in (b) which has been eliminated through the distribution of the reprojection error across the entire model.

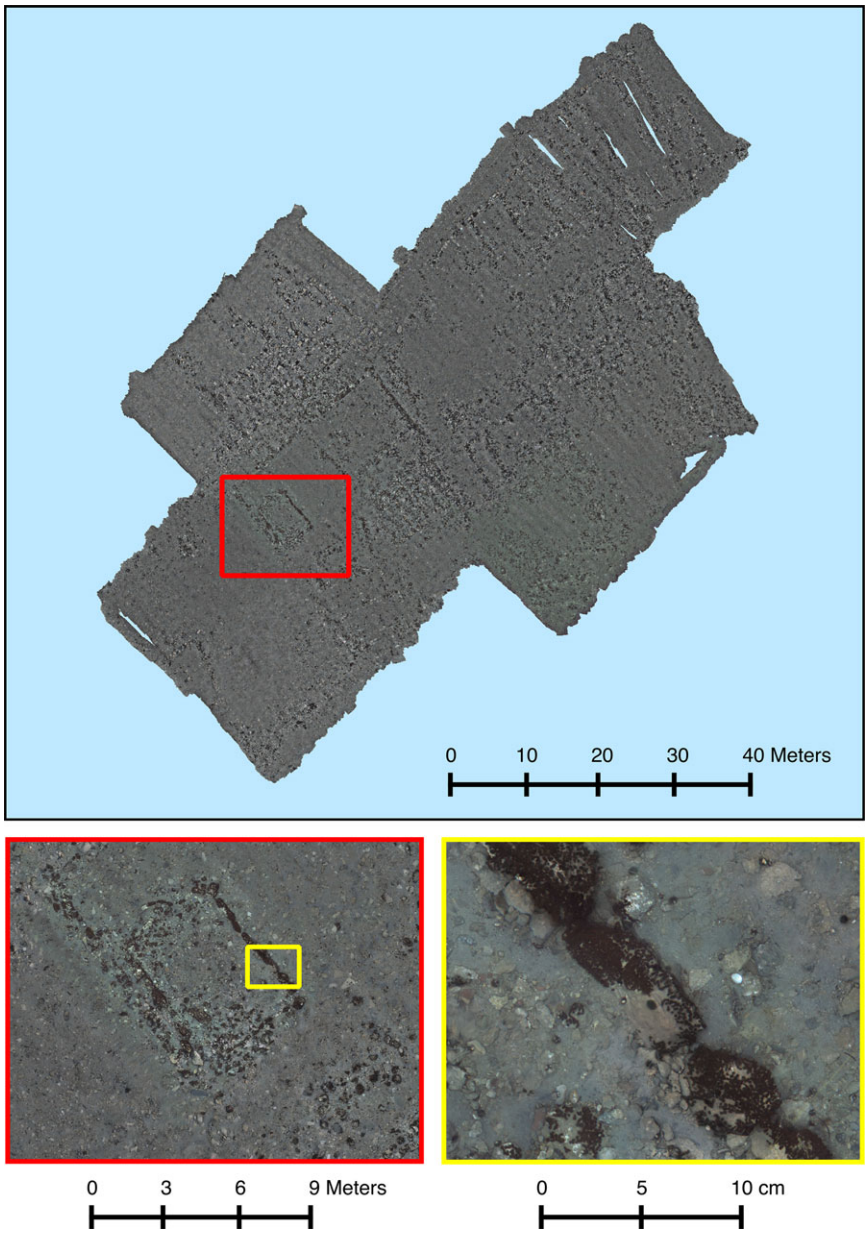

Figure 16. Optical Survey Pavlopetri Diver-rig 2010: This model is of building foundation stones from a series of buildings on the site. It is the result of the integration of over 100,000 images taken during the 2010 field season.

the difference between pixels reprojected from different independent views onto the same point on the mesh geometry. Note all images were corrected using the clustering technique described in Section 3.1. The standard deviation represents the inconsistency in both pose and structure but provides a good proxy for comparing different localization approaches. The result here is calculated over grayscale intensity only for clarity in the visualization. The individual color channel results display a similar trend.

The visual results show the proposed approach distributes the error across the mesh more broadly than our previous non-bundle-adjusted approach. Note the high deviation error that appears in the majority of the lower half of the model in Fig. 14 (b). To further bolster the claim that the proposed approach is superior in pose self-consistency, a histogram of the deviation across the mesh appears in Fig. 15. The non-bundle-adjusted approach employed previously (Mahon et al., 2011), titled SLAM, has a much

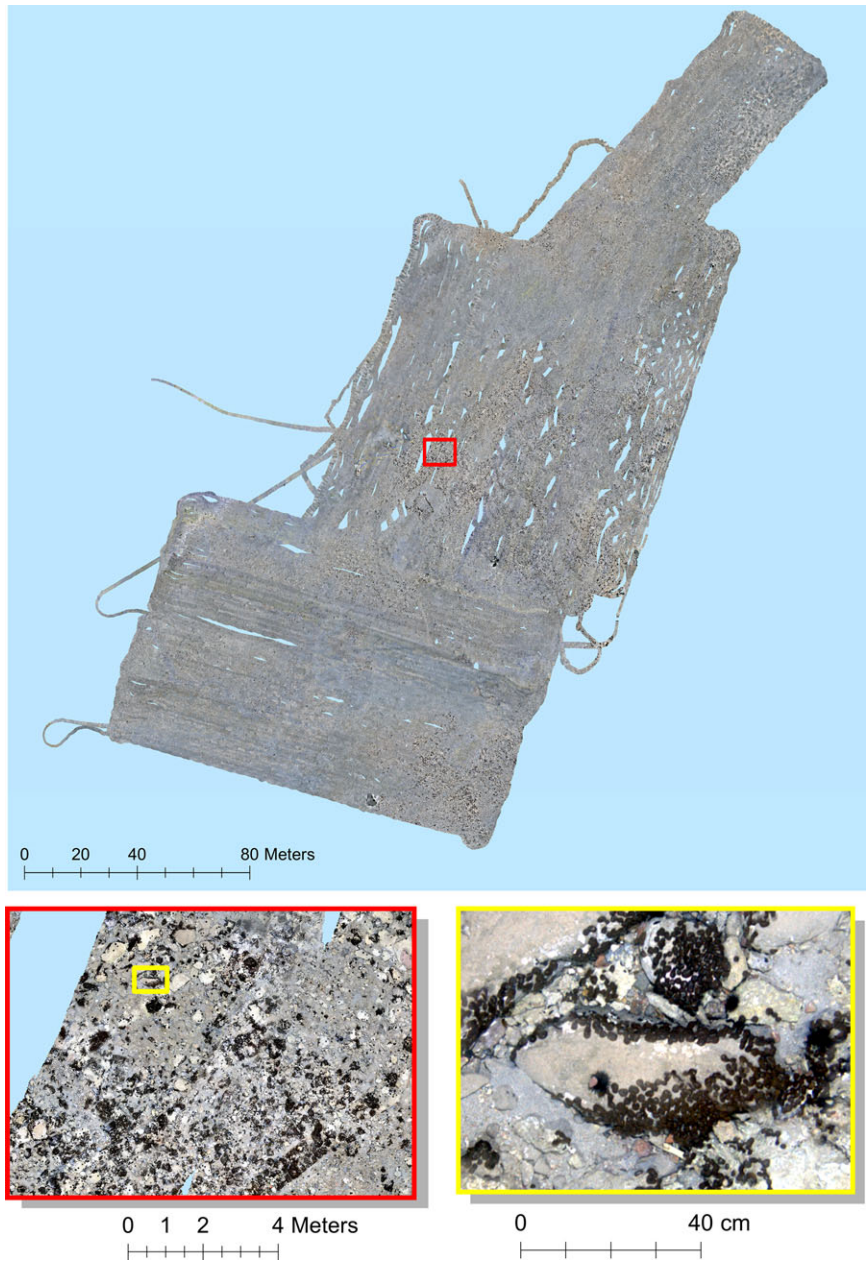

Figure 17. 3D model from the 2011 field season. This model is the result of a bundle adjustment of approximately 400,000 Iver2 gathered images into a large scale 3D model in which individual stones can be seen and mapped across the entire site.

higher frequency of high deviation pixels where the proposed approach, titled BA, has lower overall deviation and has the highest frequency of occurrence in the low deviation bins.

4.3. Broad Scale Site Map Results

The full map of the 2010 results appears in Fig. 16. Note the high-resolution that appears in the insets; this resolution gathered across a broad area enables archaeologists to review the site thoroughly without diving. The resulting model from the 2011 field season appears in Fig. 17 with depth relief appearing in Fig. 18. The bundle adjustment of a model of this size was a major thrust of this work. This model comprising approximately 400,000 images is at a scale and resolution, to the authors knowledge, previously unattempted in the underwater 3D reconstruction realm to date.

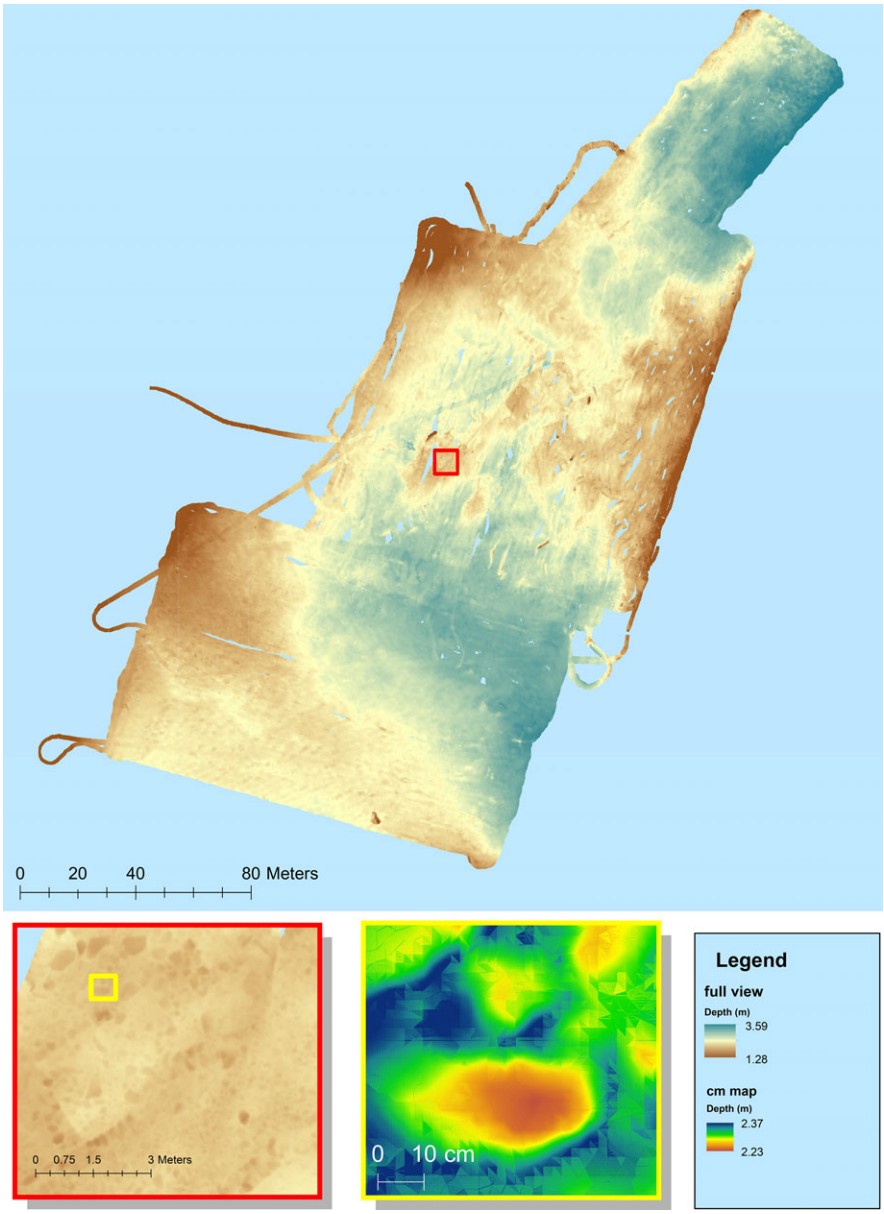

Figure 18. Relief Map from the 2011 field season. This model is the 3D depth map of Fig. 17. Note the fidelity of the model where the height map reveals stones on the centimeter scale.

5. CONCLUSIONS

In this paper, we have shown how the combination of two platforms can allow for the mapping of a large extent and in combination with frequent small-area passes can rapidly aid in the generation of archaeologically relevant data in a short period. The use of a small torpedo-class AUV for shallow-water mapping has many applications and was shown to dramatically increase the coverage area and density of images, which can be gathered in a relatively short field season.

In addition to the development of a pipeline that is is capable of processing out of core over half a million images on conventional hardware represents a leap forward in field processing for 3D reconstruction.

The ease with which large-scale reconstructions can be visualized provided a new perspective to archaeologists that are working on the site. By exploiting the diver-rig and AUV's ability to quickly map small areas, the speed at which an excavation and other archaeological activities

could proceed was greatly increased. The ability to share these reconstructions with other experts and the general public also open up new venues for collaboration and diffusion of archaeological work. Archaeologists involved in the project were excited about the ability to accurately survey submerged features quickly and, most important, to a high level of accuracy without the need for a large equipment infrastructure or a support team and considered this to be a major step forward in terms of underwater archaeological recording.

ACKNOWLEDGMENTS

This work was supported by the University of Sydney's Australian Centre for Field Robotics through the Australian Research Council (ARC) Centre of Excellence programme, funded by the ARC and the New South Wales State Government. Fieldwork was made possible through funding from the Institute of Aegean Prehistory (INSTAP), the University of Nottingham, the British School at Athens, the Hellenic Ministry of Culture, and the Municipality of Voiai. Special thanks to the Greek director of the Pavlopetri Project, Mr. E. Spondylis, and to Dr. A. Simosi, the Underwater Ephorate of the Hellenic Ministry of Culture, Professor Cathy Morgan, and Dr. Chrysanthi Gallou. We are indebted to Aggelos Mallios, Dimitris Sakellariou, and the Hellenic Centre for Marine Research (HCMR) for their generous logistics, equipment and data-gathering support. The gathering of the data would not have been possible without support from the ACFR, including Christian Lees, Michael Jakuba, Lachlan Toohey, Donald Dansereau, and Ian Mahon. We would also like to thank the field archaeology team, including Gemma Hudson, Robin Harvey, Kirsten Flemming, Peter B. Campbell, and Dr. Derek, Eugienia for marking out the survey areas and spending time underwater managing the survey guide line and poles.

REFERENCES

Agarwal, S., Mierle, K., & Others (2014). Ceres Solver. http://ceres-solver.org.

Arun, K. S., Huang, T. S., & Blostein, S. D. (1987). Least-squares fitting of two 3-D point sets. IEEE Transactions on Pattern Analysis and Machine Intelligence, 9(5), 698–700.

Ballard, R. (2007). Archaeological oceanography. In J. Wiseman & F. El-Baz, editors, Remote Sensing in Archaeology. Interdisciplinary Contributions to Archaeology, pages 479–497. New York: Springer.

Ballard, R. D., Stager, L. E., Master, D., Yoerger, D., Mindell, D., Whitcomb, L. L., Singh, H., & Piechota, D. (2002). Iron Age shipwrecks in deep water off Ashkelon, Israel. American Journal of Archaeology, 151–168.

Bay, H., Ess, A., Tuytelaars, T., & van Gool, L. (2008). Speeded-up robust features (surf). Computer Vision and Image Understanding (CVIU), 110(3), 346–359.

Bingham, B., Foley, B., Singh, H., Camilli, R., Delaporta, K., Eustice, R., Mallios, A., Mindell, D., Roman, C., & Sakellariou, D. (2010). Robotic tools for deep water archaeology: Surveying an ancient shipwreck with an autonomous underwater vehicle. Journal of Field Robotics, 27(6), 702–717.

Bowens, A. (2009). Underwater archaeology: The NAS guide to principles and practice. Oxford: Blackwell.

Bradski, G., & Kaehler, A. (2008). Learning OpenCV: Computer vision with the OpenCV library. O'Reilly.

Bryson, M., Johnson-Roberson, M., Murphy, R. J., & Bongiorno, D. (2013a). Kite aerial photography for low-cost, ultra-high spatial resolution multi-spectral mapping of intertidal landscapes. PLoS ONE, 8(9), e73550.

Bryson, M., Johnson-Roberson, M., Pizarro, O., & Williams, S. (2012). Colour-consistent structure-from-motion models using underwater imagery. In Robotics: Science and Systems (RSS), pages 1–8. MIT Press, Sydney, NSW, Australia.

Bryson, M., Johnson-Roberson, M., Pizarro, O., & Williams, S. (2013b). Automated registration for multi-year robotic surveys of marine benthic habitats. In IEEE/RSJ International Conference on Intelligent Robots and Systems. Tokyo, November 3–7, 2013.

Buchsbaum, G. (1980). A spatial processor model for object colour perception. Journal of the Franklin Institute, 310(1), 1–26.

Burt, P. J., & Adelson, E. H. (1983). A multiresolution spline with application to image mosaics. ACM Trans. Graph., 2(4), 217–236.

Capus, C. G., Banks, A. C., Coiras, E., Ruiz, I. T., Smith, C. J., & Petillot, Y. R. (2008). Data correction for visualisation and classification of sidescan SONAR imagery. IET Rdar, Sonar & Navigation, 2(3), 155–169.

Clark, J. H. (1976). Hierarchical geometric models for visible-surface algorithms. In Proc. of the 3rd Annual Conference on Computer Graphics and Interactive Techniques, pages 267–267. Philadelphia, Pennsylvania, July 14–16, 1976.

Curless, B., & Levoy, M. (1996). A volumetric method for building complex models from range images. In Proc. of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, pages 303–312. New Orleans, LA, USA, August 4–9, 1996.

Dellaert, F., & Kaess, M. (2006). Square root SAM: Simultaneous localization and mapping via square root information smoothing. International Journal of Robotics Research, 25(12), 1181–1203.

Duntley, S. Q. (1963). Light in the sea. Journal of the Optical Society of America, 53(2), 214–233.

Foley, B. P., Dellaporta, K., Sakellariou, D., Bingham, B. S., Camilli, R., Eustice, R. M., Evagelistis, D., Ferrini, V. L., Katsaros, K., & Kourkoumelis, D. (2009). The 2005 Chios ancient shipwreck survey: New methods for underwater archaeology. Hesperia, 78, 269–305.

Green, J. (1990). Maritime archaeology. London: Elsevier/Academic Press.

Harding, A. F., Cadogan, G., & Howell, R. (1969). Pavlopetri, an underwater Bronze Age town in Laconia. Annual of the British School at Athens, 64, 112–142.

Henderson, J., Gallou, C., Flemming, N., & Spondylis, E. (2011). The Pavlopetri underwater archaeology project: Investigating an ancient submerged town. Underwater Archaeology and the Submerged Prehistory of Europe, pages 207–218. Wiley/Blackwell, Hoboken, New Jersey.

Henderson, J., Pizarro, O., Johnson-Roberson, M., & Mahon, I. (2013). Mapping submerged archaeological sites using stereo-vision photogrammetry. International Journal of Nautical Archaeology, 42(2), 243–256.

Henderson, J. C., & Burgess, C. (1996). Close contour survey of submerged sites to create data terrain models. International Journal of Nautical Archaeology, 25(3), 250–256.

Huber, P. J. (1964). Robust Estimation of a Location Parameter. Annals of Mathematics Statistics, 35(1), 73–101.

Johnson-Roberson, M., Bryson, M., Douillard, B., Pizarro, O., & Williams, S. B. (2013). Out-of-core efficient blending for underwater georeferenced textured 3d maps. In IEEE Computing for Geospatial Research and Application (COM. Geo), 2013 Fourth International Conference on, pages 8–15. San Jose, CA, July 22–24, 2013.

Johnson-Roberson, M., Pizarro, O., Williams, S. B., & Mahon, I. (2010). Generation and visualization of large-scale t reconstructions from underwater robotic surveys. Journal of Field Robotics, 27(1), 21–51.

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2), 91–110.

Ludvigsen, M., Sortland, B., Johnsen, G., & Singh, H. (2007). Applications of geo-referenced underwater photo mosaics in marine biology and archaeology. Oceanography, 20, 140–149.

Mahon, I., Pizarro, O., Johnson-Roberson, M., Friedman, A., Williams, S. B., & Henderson, J. C. (2011). Visually mapping Pavlopetri–the world's oldest submerged town. In IEEE International Conference on Robotics and Automation. Shanghai, China, May 9–13, 2011.

Maimone, M., Cheng, Y., & Matthies, L. (2007). Two years of visual odometry on the Mars exploration rovers. Journal of Field Robotics, 24(3), 169–186.

Martin, C. J. M., & Martin, E. A. (2002). An underwater photomosaic technique using Adobe Photoshop. International Journal of Nautical Archaeology, 31(1), 137–147.

Mitchell, N. C., & Somers, M. L. (1989). Quantitative backscatter measurements with a long-range side-scan sonar. IEEE Journal of Oceanic Engineering, 14(4), 368–374.

Pizarro, O., & Singh, H. (2003). Toward large-area mosaicing for underwater scientific applications. IEEE Journal of Oceanic Engineering, 28(4), 651–672.

Royal, J. G. (2012). Illyrian Coastal Exploration Program (2007–2009): The Roman and Late Roman finds and their contexts. American Journal of Archaeology, 116(3), 405–460.

Rzhanov, Y., Cutter, G. R., & Huff, L. (2001). Sensor-assisted video mosaicing for seafloor mapping. In Proc. International Conference on Image Processing, pages 411–414. Thessaloniki, Greece, October 7–10, 2001.

Sakellariou, D. (2007). Remote sensing techniques in the search for ancient shipwrecks: How to distinguish a wreck from a rock in geophysical recordings. Bulletin of the Geological Society of Greece, 37(4), 1845–1856.

Sedlazeck, A., Koser, K., & Koch, R. (2009). 3D reconstruction based on underwater video from Rov Kiel 6000 considering underwater imaging conditions. In OCEANS 2009-EUROPE, pages 1–10. IEEE. Bremen, Germany, May 11–14, 2009.

Steinberg, D., Friedman, A., Pizarro, O., & Williams, S. B. (2011). A Bayesian nonparametric approach to clustering data from underwater robotic surveys. In International Symposium on Robotics Research, Flagstaff, Arizona, USA, August 28–1 September, 2011.

Steinberg, D., Pizarro, O., & Williams, S. (2013). Synergistic clustering of image and segment descriptors for unsupervised scene understanding. In Computer Vision (ICCV), 2013 IEEE International Conference on, pages 3463–3470. Sydney, Australia, December 1–8, 2013.

Thrun, S., Burgard, W., & Fox, D. (2005). Probabilistic robotics (Intelligent Robotics and Autonomous Agents). Cambridge, MA: MIT Press.

Triggs, B., McLauchlan, P. F., Hartley, R. I., & Fitzgibbon, A. W. (2000). Bundle adjustment–a modern synthesis. In Proceedings of the International Workshop on Vision Algorithms: Theory and Practice, ICCV '99, pages 298–372, London, UK, UK. Springer-Verlag. Corfu, Greece, September 21–22, 1999.

Umeyama, S. (1991). Least-squares estimation of transformation parameters between two point patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(4), 376–380.

Williams, S., Pizarro, O., Jakuba, M. V., Johnson, C. R., Barrett, N. S., Babcock, R. C., Kendrick, G. A., Steinberg, P. D., Heyward, A. J., Doherty, P., Mahon, I., Johnson-Roberson, M., Steinberg, D., & Friedman, A. (2012). Monitoring of benthic reference sites using an autonomous underwater vehicle. IEEE Robotics and Automation Magazine, 19(1), 73–84.