Surf Zone Characterization Using a Small Quadcopter: Technical Issues and Procedures

Rob Holman, Kate Brodie and Nick Spore

0.0 Abstract

We explore the potential for using a small Unmanned Aerial Vehicle (UAV) quadcopter to collect long-dwell imagery of the nearshore from which important measurements can be made at low cost and with flexibility. This work extends existing topographic imaging approaches that rely on having plentiful ground control spread across the image, to the nearshore case where the bulk of the image is water with no control point and vehicle metadata must be used. The UAV autopilot was found to be capable of excellent station-keeping with positional errors of 0.20 and 0.53 m (horizontal and vertical) and viewing angle errors of 0.25° (tilt and roll) and 0.38° (azimuth). The ground position of imaged objects could be found to 0.21 m accuracy. Metadata returned by the UAV on camera position was accurate to 5 m, and the camera roll could be assumed to be 0°, reducing the ground control requirements to two, or even one location. Even under this extreme simplification, ground position errors averaged only 10 m but were worst for cases when only control points near to the UAV were used. A model for the visual contrast of waves when viewed from different angles found that large tilts are important but, in contrast to theory, that there was little dependence on the viewing azimuthal angle. Derived Argus products agreed well with the same products collected using a traditional fixed Argus station. UAVs appear to be a very promising alternate to fixed camera systems if limited duration sampling is adequate.

Index Terms: Nearshore, remote sensing, Argus, UAV, surf zone.

1.0 Introduction

Rising sea level and increasing population pressure on ocean coasts have focused the need for improved understanding of coastal dynamics and the ability to make predictions on both short and long time scales. Success will require good understanding of the relevant physics implemented through models, but also an improved ability to make relevant measurements at low cost and over long time scales that can serve as the input data for these models. This has been the province of nearshore remote sensing. While this can be accomplished using a number of sensor types including optical, infrared, radar and LIDAR, optical methods have been the most common due to their low cost and ease of logistics as well as their natural appeal of optical data that is so familiar to humans. Holman and Haller [2012] provide a discussion of the various modes of nearshore remote sensing.

Within the optical domain, a number of systems have been developed including Cam-Era (https://www.niwa.co.nz/our-services/online-services/cam-era), the COSMOS system [*Taborda and Silva*, 2012] and many others, with the earliest system being the Argus system developed at the Coastal Imaging Lab at Oregon State University beginning in 1986 [*Holman and Stanley*, 2007]. Using fixed cameras located on high vantage points such as lighthouses and tall coastal buildings, Argus samples in both image product as well as time series mode on a user-defined schedule (usually every daylight hour). Because viewing directions are fixed, image geometries need only be found once and data collected over time can be directly compared without complication.

In many cases where coastal measurements are needed, no high vantage point is available. Additionally, there is often a need for video measurements for shorter periods of time so that the complication of a full Argus station installation is not worthwhile (or may be unavailable). In these cases, the recent development of small Unmanned Aerial Systems (sUAS, or drones or UAVs, for short) offers an option that is becoming increasingly attractive. Surprisingly capable systems can purchased for around $1000 (US) and can be flown safely even by non-specialists (all users should be aware of and practice safe and legal procedures). It is of interest to many, and the goal of this and other papers, then to determine the feasibility of using UAV data as an alternate to traditional Argus sampling. Additionally, UAVs offer a choice of viewing angles that is not available to a fixed Argus station. We wish to examine the physical basis for choosing viewing angles to optimize wave contrast.

A number of papers have already addressed components of this problem. Pennucci et al [2007] studied the potential of a small but standard military fixed wing system, the Raven B manufactured by AeroVironment, finding that image quality was good, even with the limited cameras available at that time, but that image navigation required good ground point control and that viewing dwell times were usually less than a minute. The application of Argus algorithms for the Raven was studied further by Holman et al [*Holman et al.*, 2011], again pointing out the difficulty of achieving significant on-target dwell with a fixed wing system using a fixed camera view. In addition, these systems are prohibitively expensive for non-military applications.

The situation improved dramatically with the recent marketing of small hobby-level systems, particularly small quadcopters with excellent onboard stabilization both of the camera and flight. Brouwer et al [2014] were among the first to test multi-copters for nearshore applications, using an Altura AT6 and a Y6 hexacopter in an alternating flight pattern to achieve near continuous coverage of rip current plumes over many hours. They found these to be very useful in dye tracer studies and that they featured a station-keeping accuracy of a few meters and ground accuracy of imaged features, assuming good availability of visible ground control points, of order 1 m.

Turner et al [2016] demonstrated the use of a good fixed-wing UAV (the SenseFly eBee-RTK) as a solution to the problem of measuring coastal topography, returning dense 3D point clouds of sub-aerial topography over several kilometers of beach, computed using the Pix4D structure-from-motion (SfM) software package and surveyed ground control points. Accuracy compared to GPS surveys was found to be roughly 0.07 m, comparable to the expected accuracy of the GPS "ground truth" data. The eBee is considerably more expensive ($25,000 at the time of writing) due partly to the RTK GPS system that removes the need for as much ground control.

In fact, topographic point cloud estimation has seen a rapid increase in activity largely due to the availability of lower cost UAV systems and turn-key commercial software packages like those available from Pix4D and Agisoft. Toth [2015] provides a very good review of the topographic estimation problem for fixed ground scenes using three commercial (unnamed) software packages for the structure-from-motion calculations and low cost quadcopters. These scenes are characterized by a rich set of surveyed Ground Control Points (GCPs) spread across the field of view and typically thousands of automatically-generated point matches between image pair that are used for the aerial triangulation used to build the 3D point cloud.

The structure from motion methods above rely on having many viewing angles (images) to allow stereo solution, and having many control points (GCPs) spread across the field of view to align the point cloud to a world coordinate system. These requirements cannot be met for many nearshore scenes. Typically nearshore imagery is intended for sampling waves and currents and features mostly water with only a small edge of ground view. Similarly, the Argus sampling approaches that we wish to extend to UAV data commonly exploit fixed-view time series data sampled at 2 Hz over a 17-minute collection so do not include the camera motion that is needed for stereo solution of SfM. Thus these commercial packages cannot be used for UAV sampling strategies that try to mimic Argus data collection and more traditional approaches must be used.

The goal of this paper is to investigate the potential for using UAVs as a substitute for or supplement to fixed Argus camera data collection, and the potential to use all of the types of data collections and analyses used by Argus for making measurements. This will include determination of the quality of imagery and image stabilization as well as methods available for stabilizing and georectifying imagery collected with minimal ground control as well as an investigation of the sensitivity of imaged wave contrast to the choice of viewing angles. The results will be a set of recommended procedures, with accuracies, that can be used.

The analysis will be focused on one specific UAV and results will necessarily be specific to that platform. However, the methodology and performance measures can and should be applied to any new platform as a precursor to scientific sampling.

Section 2 of this paper will discuss the methods used including the characteristics of the UAV used for the study, the photogrammetric methods used, and the nature of the bathyDuck field experiment used to test these methods. Section 3 will cover the results including those describing the UAV flight and stability characteristics, viewing angles sensitivities and the available returned products including the feasibility of running cBathy analysis for sub-aqueous bathymetry. Section 4 will discuss concepts of operations and comparison of the tested UAV with a previous generation but very popular alternate. This will be followed by conclusions.

2.0 Methods

2.1 Selected UAV Platform

Many platforms are available for making airborne measurements (for example, see Toth et al [2015]). Because we required the capability to collect 17-minute (1024 s) time series data we selected a multi-rotor helicopter, opting for the very popular DJI Phantom 3 Professional (P3P) quadcopter due to its low price, high quality, ease of use, and excellent imagery (Figure 1). The P3P weighs 1.3 kg, can transit at speeds up to 16 m/s and is quoted as being able to hover with a few meter positional accuracy for flight times that can exceed 20 minutes.

The camera features a 4000 by 3000 (12.4 Mpixel) 1/2.3" Sony sensor chip with an 82° horizontal field of view (quoted as 94° on the manufacturer's web page). The camera was tested in two of the five collection modes, a '4K' video mode collected at 30 Hz with 3840 by 2160 resolution, and a 4000 by 2250 (16:9 high definition aspect ratio) snapshot mode. 4K videos rapidly become very large and are automatically broken into roughly 3 GB parts for onboard storage.

Temporal control is based on received GPS signals so is assumed to be very accurate. The camera lens was calibrated using the Caltech Lens Calibration software package (http://www.vision.caltech.edu/bouguetj/calib_doc/), correcting to sixth order for radial distortion and second order in the tangential direction.

2.2 Photogrammetric Methods

In this section, we discuss the photogrammetric methods available for georectification. These are embedded in SfM software packages but must be explicitly handled for our case of sampling from a fixed location since no motion is available to exploit.

The connection between the locations of objects in the world and their corresponding location in an image is described by photogrammetric relationships. The form of these relationships that we use is taken from Hartley and Zisserman [2003] and uses the concept of homogeneous coordinates. It is well described in several references but is summarized here for completeness and since many of the implementation details require this knowledge.

By convention, objects in the world are described by the 3D coordinates, [x, y, z] (cross-shore, longshore, vertical), while their image locations are described by the 2D coordinates, [U, V] (both are right hand coordinate systems). In a homogenous formulation, the two are related through a 3 by 4 projective transformation matrix, P, such that

$$\begin{bmatrix} U \\ V \\ 1 \end{bmatrix} = P \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \qquad (1)$$

The normal 2 and 3D vectors are each augmented by an additional coordinate, set to the value of 1. Thus, for any particular world location, if P is known, the image location is found by the multiplication in equation (1). In homogeneous coordinates, the answer on the left is considered to be known to a multiplicative constant. That means that the literal product of the multiplication in (1) will yield a non-unitary last component, but this is logically equivalent to what you would get by dividing by the last value, in which case the first two components are the image coordinates of the object. Thus, computation of image coordinates requires first the multiplication, then the normalization to make the last element equal to 1. There are many benefits to make up for the inconvenience of the second step.

The projective matrix is composed of three factor matrices,

$$P = KR\left[\begin{array}{c|c} I & -C \end{array}\right] \qquad (2)$$

K contains the intrinsic parameters of the camera, those that convert from angle away from the center of view into camera coordinates. R is the rotation matrix describing the 3D viewing direction of the camera compared to the world coordinates system. The final bracketed term is a 3 by 3 identity matrix, I, augmented by C, a 3 by 1 vector of the camera location in world coordinates. Taking the multiplication (equation 1) in steps, first multiplying the bracketed term in (2) by the object world coordinates causes subtraction of the camera location from the object location, effectively putting the object in camera-centric coordinates. Then multiplying by the rotation matrix rotates into directions relative to the camera look direction. Finally, multiplying by K, the intrinsic matrix, converts into pixel units for the particular lens and sensing chip.

The intrinsic parameter matrix, K, is a function of the camera lens and chip and is not a function of the specific installation, i.e. the camera location and viewing angles. As a consequence, the parameters in K are found during a lens calibration prior to camera installation. We use the excellent Caltech calibration package (http://www.vision.caltech.edu/bouguetj/calib_doc/). The form of K is

$$K = \begin{bmatrix} f_U & s & U_0 \\ 0 & f_V & V_0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (3)$$

Here f_U and f_V are the focal lengths in the U and V directions, expressed in pixels, U_0 and V_0 are the coordinates of the principal point (geometric image center), and s is the image skewness (cosine of the angle between the U and V axes) and is assumed to be 0.0. K has 5 degrees of freedom (DOF) with values returned during the calibration process. Because the number of degrees of freedom will be important to the following discussions, we will use numbers rather than words to enumerate them.

Note that the calibration process also computes estimates of lens distortion parameters, used to convert between image locations from the camera and those that would have been returned from a perfect camera with no lens distortion. Some cameras such as those with fish eye lenses exhibit severe barrel distortion, for example a highly curved horizon that must be corrected for. But even the fairly accurate P3P lens requires calibration and distortion removal. This process is always used but is not described further in the discussion below (see the Caltech toolbox).

The rotation matrix, R, represents the 3D rotation between world and camera coordinates. There are 3 degrees of freedom, the azimuth (taken here as the compass-like rotation clockwise from the positive y-axis), the tilt (zero at nadir, rising to 90° at the horizon), and roll (rotation about the look direction, positive in the counter-clockwise direction as viewed from the camera). The details can be found on page 612 in Wolf [1983].

Finally, the camera location, C, has 3 degrees of freedom, its 3D world location. Thus, there are 11 total unknowns of which 5 can be solved during calibration and 6 must be found after camera placement (the 3 coordinates of the camera location and the 3 rotation angles). In general, these values are found using GCPs, points whose world coordinates are known by survey and whose image coordinates can be accurately digitized from an image. Combining equations (1) and (2) and applying these to a set of such points, the only unknowns will be the 6 camera parameters so these can be found by a standard nonlinear solver (comparing measured and predicted image coordinates for a guess at the 6 unknowns then searching for their optimum values that minimize the squares of their differences).

Since there are 6 unknowns, we need at least 6 knowns for a solution. Each control point contributes 2 values (U and V coordinates) so at least three points are needed. We prefer to be over-determined so will use at least four points in the following tests. For terrestrial applications it is typically easy to find or place an abundance of GCPs throughout the view to allow solution of camera extrinsic geometries, the heart of SfM algorithms. However surf zone images usually contain only a minimum amount of land by design, so GCP options are often limited and poorly distributed over the image, often lying in a line along the dune crest, a configuration that makes the inverse solution ill-posed. For these cases, common for nearshore studies, we must rely on alternate sources of information to reduce the number of degrees of freedom and the requirements on GCP layout.

It is rare to find sufficiently accurate information of the azimuth and tilt of an airborne camera so these variables almost always must be solved for. However the camera location is often available in the imagery, based on an onboard GPS system, and can be extracted, for example by using exiftool or other image information packages. Vertical position is also often returned in the image metadata and could be used if no better GCPs are available. However, it is usually less accurate than horizontal position data. For example, altitude may be expressed relative to the takeoff point, rather than in a global coordinate system or it may be a low-quality uncorrected GPS measurement. Finally, it is reasonable to assume for a good stabilized gimbal as on the Phantom 3 that roll is stable and can perhaps taken as equal to 0° for a reasonable approximation. Thus it is possible to reduce to as low as two unknowns which can be solved in a least squares sense with just two GCPs anywhere on the image or in a non-least squares sense with just a single GCP. The relative accuracy of these alternate assumptions will be tested below.

2.3 Field Methods

The BathyDuck field experiment took place from 28 September to 1 November, 2015. One component was a set of test flights using small commercial off-the-shelf (COTS) quadcopters to determine their capability for making usable optical measurement of the nearshore and to develop appropriate concepts of operations (conops) to simplify and optimize sampling.

The site is the home of the US Army Corps of Engineers Field Research Facility (FRF), the location of many previous nearshore community experiments. The beach (see example snapshot, Figure 2) is typical of US East Coast barrier islands and is an intermediate beach with one to two active sand bars and a relatively steep foreshore. Wave data were collected from the offshore 26 m Waverider and beach surveys were carried out on 01, 04, 09, 19, 28, and 30 October using the FRF LARC and CRAB and are considered to be accurate to 5 cm [*Birkemeier and Mason*, 1984]. Wind data were collected by a pier-end Met station. Winds picked up soon after the beginning of the experiment, being marginally flyable on October 1 (11.51 m/s!) before becoming un-flyable for the next six days. Winds dropped during the night of October 6 and data collections resumed.

Ground Control Points consisted of 1 m square black-and-white checkerboard targets that were distributed across the scene and surveyed into the local coordinate system to an accuracy of several centimeters using RTK-GPS. Due to the presence of the pier, GCPs could be deployed in a way that was not collinear (had spread in two image dimensions), allowing full 6 degree of freedom solutions. These solutions were considered to be the best available and the degraded solution methods mentioned above were compared to these full solutions for performance appraisal. GCP locations were manually selected by clicking in the images. Replicate tests of ten digitizations of the four points show that this processes is reproducible to within a standard deviation of 1.5 pixels or better than 0.20 m for the four locations shown in Figure 2.

Since the camera aim drifts slightly between frames, the geometry must be found for each subsequent frame individually, a tedious process if each frame required manual digitization. Instead, we identify image features that can be automatically recognized by, in this case, being brighter than its surroundings within a small search window. After the GCPs have been digitized and the geometry of the first frame has been found, a number (commonly four) of such search window and intensity thresholds are identified by the user. The center of mass of the thresholded target window is then found to a sub-pixel accuracy and assigned a virtual world position that is equivalent to the center of mass U,V coordinates and a user-assigned vertical (z) location (inverting equation (1) but with a known z). Thus, in each subsequent frame, the location of these virtual GCPs is found automatically and the 6 degree of freedom geometry of the frame found and saved. These data provide us with a time series of variations in the camera position and viewing angles from which we can assess the stability of the camera view.

3.0 Results

In this section we will assess several aspects of the performance of the P3P drone. First, we will use manual digitization of GCPS in individual snapshots to test the repeatability of the GCP inversion process, the accuracy of a typical 6 DOF fit and the loss of accuracy of reduced DOF fits where we use various pieces of vehicle metadata as part of the 6 unknowns. Second, we will use 17-minute time series data from UAVs in flight at fixed locations to assess the stability of the vehicle and the gimbal system. Third, we will examine the dependence of wave contrast on the viewing angles tilt and azimuth to better understand optimum sampling strategies. Finally, we will look at several usual Argus products that we can produce from UAV data and compare to the same products collected by a traditional fixed Argus station.

3.1 Accuracy of Reduced DOF Solutions.

To test the basic solution methods, a set of five representative snapshots were selected. For each, we had available the exiftool metadata of GPS latitude-longitude, converted to local coordinates and the estimated elevation. Four GCPs were manually digitized for each snapshot, each with three replicates, and the 6 DOF solutions found for each case. The three replicates showed that on average the solutions for the camera location were consistent within 0.21 m (standard deviation), approximately the accuracy of manually digitized GCP input for this representative camera height and viewing angles (Table 1, last three columns).

If we take these values as a reasonable approximation of truth, we can compare them to the returned exif image data of GPS latitude and longitude, converted to the local coordinate system, and elevation. The mean differences were 4.6 and 3.1 m in x and y (Table 1) while the mean deviation for the vertical coordinate was 10.6 m. Vertical meta data levels are estimated relative to the takeoff elevation, 5.08 m for these tests so there remains an average error of ~5 m in the vertical. Thus, it appears that the metadata locations can be used with an expected error of about 5 m (standard deviation).

The standard deviation among the replicates of azimuth, tilt and roll were 0.06, 0.11 and 0.1 degrees, respectively, so typically about 0.1 degrees. The measure of interest is the roll, and particularly whether we can assume that the roll can be considered fixed and need not be solved for. The global mean roll was -0.40° with a standard deviation of 0.58°. The P3P has the capability to adjust the roll to remove this error, but this was not done prior to flight.

Since we wish to reduce the number of unknowns by substituting metadata, we next tested the consequences of different assumptions on the accuracy of geo-rectified products. This was tested as follows. If the camera location, $[x_c, y_c, z_c]$, is assumed to be known, only the three viewing angles remain unknown and can be solved for using only two GCP locations. Given that we have already digitized four GCP locations for each of the five test images (e.g. Figure 2), we can use these in pairs or triplets to solve for the viewing angles. We can then compute the equivalent ground GCP locations using the digitized UV coordinates of each and the solved for camera viewing geometries and assuming the vertical coordinate of each GCP is known. We can then find the ground distance error between this estimated GCP locations and the surveyed location. The results depend on the specific locations of the selected GCPs, both those that are and aren't used in the solution.

This test results in 120 realizations, six permutations of two GCPs out of the four available, times four GCP locations per image, times five images. Table 2 contains the error statistics for various combinations of assumed knowns and unknowns and for cases where the solution was based on the least squares fit to one, two, or three, GCPs. For comparison, the full 6 DOF solution computed using all four GCPs had a mean error of 0.9 m with a worst case among the five images of 4.9 m. The mean ground distance error for the case of a known camera location and two GCP solutions was 10.2 m with a median, 95th percentile and maximum error of 6.9, 30.6 and 51.8 m, respectively. In other words, much of the misplacement error within the field of view will be around 10 m, but errors can exceed this substantially for certain choices of GCP and viewing points.

Several generalizations can be made. Not surprisingly, the use of more than the minimum numbers of degrees of freedom reduces error. It is apparent that the inclusion of roll as a known also reduces solution error, despite the assumed value of 0.0° being slightly wrong. Thus if only three GCPs can be seen and if they are collinear (say along a dune crest), least squares solutions can be made possible by assuming a value for roll (usually 0°). The other conclusion of this analysis is that the worst-case errors were always associated with the use of near-field GCPs, i.e GCPs that are close to the camera. A camera positional error of 5 m introduces a larger estimated tilt error for a nearby GCP than for one more distant. This then becomes magnified for ground distance error for distant imaged points. A suggested conops then might be to choose a camera placement such that important control points are not close by (perhaps by flying offshore and viewing landward).

The minimum solution option, and one that might be popular, is to assume knowledge of the camera position and the roll, leaving only the two unknowns of azimuth and tilt, and to then use only a single GCP (two knowns) for the solution. This is even determined so can be solved, but not in a least squares sense (so no error estimates are returned). This case was tested for each of the five images picking each of the four GCPs in turn and computing position errors for the remaining GCPs. The errors (Table 2) average 10 m but can become large, especially if the chosen GCP is close to the camera, amplifying errors in camera position. Still, reasonable products can be computed, for example the rectified snapshot in Figure 3. While the one GCP solution (left panel) looks convincingly like the full solution found using all four GCPS (right panel), differences are clear such as the slightly tipped orientation of the pier with the 2 DOF rectification as well as the roughly 30 m longshore displacement of the CRAB surveying vehicle near x = 275, y = 900 (the elongated triangle). It is again important to remember that the solution will be perfect in the area of the selected GCP and will degrade with distance away. Thus choosing a camera location such that the region of interest and the GCP are a similar distance away will reduce error.

3.2 P3P Platform Characteristics and Conops

The DJI Phantom 3 is a low cost, capable and easy-to-fly platform. Flights were designed to maximize dwell, the on-station video recording length, since this is at the heart of many of the Argus signal processing methods that we hope to use. Thus the typical conops was to fly directly to a good location and view, collect a single snapshot that would record the exif metadata, then turn on video recording until the low battery warning sounded, then return home.

The choice of a good location and view depends on several factors. Spatially, there is a tradeoff between the required pixel resolution and horizontal coverage since the number of pixels is fixed by the sensor as 3840 by 2160 and can be distributed with a wide spacing for a large coverage area, or with a small spacing (high resolution) and a smaller coverage area. Figure 4, a resolution map [*Holman and Stanley*, 2007] for the image used in Figure 3 (flown at 79.6 m altitude), shows that the typical mid-image pixel ground resolution of 30 by 60 cm in cross-range and range directed views respectively. This resolution is certainly sufficient for most purposes, so we find that cameras can be effectively flown at an altitude of ~100 m, below US legal limits of 122 m, and with a typical tilt of 50 to 75°tilt (the tilt of the image in Figure 2 is 68°). The dependence of wave contrast on viewing angles is discussed in the next section.

The P3P is surprisingly robust to weather, comfortably maintaining station in up to 10 m/s (20 knot) winds. While flying in rain is discouraged, several flights in light drizzle were successful as long as the camera was pointed roughly downwind to keep rain off of the lens. This occasionally required a takeoff facing downwind followed by flying backwards to the desired location. Full flight durations varied between 19 and 21 minutes for stationary (parked at altitude) data collections although only 16-17 minutes of that time was on-station video recording.

The P3P automatically maintains its three-dimensional position in the absence of operator control movements using an autopilot, thus can be easily "parked" for data collection. Figure 5 shows an example of its station-keeping ability for the 560 s data run from 10/07/15 at 1530 EDT, calculated using the full 6 dof solution. The standard deviation of the x, y, and z components of camera location were 0.21, 0.31 and 0.30 m, respectively, supporting the assumption that the camera position can be considered fixed in time. The standard deviation of the three viewing angles was 0.48°, 0.18° and 0.28° for azimuth, tilt and roll. Figure 5 shows that the tilt is the best constrained and that the small variations in viewing angles tend to be at time scales of a minute or more. The mean roll for this run was 0.57°, not zero as was hoped, but close.

Table 3 shows the stability statistics for 10 data runs of various lengths. The camera position variability is consistently much better than 1 m with the vertical variability about twice the magnitude of the horizontal component. The viewing angles vary over several tenths of a degree (standard deviation) with azimuth the least stable, presumably because it is compass-based. The third run was notably noisier than the others with significant high frequency variability. No explanation is known.

3.3 Sampling Waves

Often the goal is to specifically sample the wave field, for example for cBathy analysis (described in section 3.4). The optical contrast of imaged waves depends on the viewing angles, both in terms of the tilt, τ, as well as the azimuthal angle relative to the wave direction, α, and potentially to the direction of solar illumination. We wish to understand the impact of choices in viewing angles on the magnitudes of the resulting measured signals. Theoretical dependencies are discussed below followed by a comparison to measured test results. The theory below generally follows Walker [1994].

The light that reaches the camera from any imaged sea surface location consists of ambient skylight that has been reflected from the sloped ocean surface, and upwelled radiation, light that has entered the water, scattered off particles or the bottom and reflected back through the ocean surface to the camera. The primary optical signature of waves is due to variations in the reflected component and particularly its dependence on the varying sea surface slope associated with waves.

In turn, this reflected light varies because the source skydome illumination, the brightness of the blue sky above, is not uniform and, more importantly, because the reflection coefficient of that light back to the camera depends of the angle of incidence, γ, of the light with the normal to the sea surface. For cases of skydome illumination away from the sun and not very close to the horizon, or for overcast skies, we can neglect skydome variations and consider only the variations of intensity due to the varying reflection coefficient, i.e. $I_c = R^*I_s$, where I_s is the skydome radiance, R is the reflection coefficient, and I_c is the radiance seen by the camera. R is described by the well-known Fresnel reflection coefficient [e.g. *Walker*, 1994]

$$R(\gamma) = \frac{1}{2}\left[\frac{\sin^2(\gamma-\gamma')}{\sin^2(\gamma+\gamma')} + \frac{\tan^2(\gamma-\gamma')}{\tan^2(\gamma+\gamma')}\right] \qquad (4)$$

where γ′ is the angle of refraction (the angle of light ray propagation after it has entered the water), found by Snell's law as $\sin(\gamma) = 1.34 \sin(\gamma')$. Equation (4) is inconvenient due to the two angles but can be replaced by an empirical simplification expressed only in γ,

$$R(\gamma) = R_0 + (1 - R_0)\exp\left(\lambda\left[\gamma - \frac{\pi}{2}\right]\right) \tag{5}$$

where R_0 is the reflection coefficient at nadir (0° incidence) and is known to be 0.02 from equation (4). The best-fit value for λ was 6.20. Figure 6 shows this dependence for the Fresnel equation and for this approximation.

The angle of incidence can be found by the dot product between the sea surface unit normal vector, $\hat{r}_n$, and the camera unit normal, $\hat{r}_c$, where

$$\hat{r}_n = \begin{bmatrix} -\frac{\partial \eta}{\partial x} & -\frac{\partial \eta}{\partial y} & \sqrt{1 - \left(\frac{\partial \eta}{\partial x}\right)^2 + \left(\frac{\partial \eta}{\partial y}\right)^2} \end{bmatrix} \tag{6}$$

and

$$\hat{r}_c = \begin{bmatrix} -\sin(\tau)\cos(\alpha) & -\sin(\tau)\sin(\alpha) & \cos(\tau) \end{bmatrix}. \tag{7}$$

Since α is the azimuthal angle between the camera and the wave propagation direction, we can temporarily define the x-axis as the direction of wave propagation so the longshore gradient of sea surface slope will always be 0. Using the dot product formulation, we find the incidence angle varies as

$$\gamma = \cos^{-1}\left(\frac{\partial\eta}{\partial x}\sin(\tau)\cos(\alpha) + \cos(\tau)\sqrt{1-\left(\frac{\partial\eta}{\partial x}\right)^2}\right) \qquad (8)$$

Thus incoming waves yield an oscillating sea surface normal aligned in the temporary x plane that gives reflection variations and the combination of equations (5) and (8) determine the variance of observed optical intensity. We wish to know the dependence of this variance on viewing angles, a relationship known as the Modulation Transfer Function, Γ_F (the ratio of measured optical variance to ocean wave variance). The subscript F denotes that this is the MTF associated only with Fresnel reflection. Expressing the wave in the normal way as

$$\eta = a\cos(kx + \sigma t) \qquad (9)$$

where k is the wavenumber (equals 2π divided by the wavelength), σ is the radial frequency (equals 2π divided by the period) and wave propagation is shoreward, then the variance over a wave period is $a^2/2$. The equivalent variance of optical intensity can be found digitally by integration over a wave period using equations

(5) and (8). An example for an 8 s wave is shown in Figure 7 expressed as $\log_{10}$ of the variance of the reflected intensity signal.

The observed signal drops dramatically as tilt goes toward nadir, for example dropping four orders of magnitude (two orders in standard deviation space) for a tilt of 50° and eight orders of magnitude near nadir where the sensitivity of R to changing incidence angle in Figure 6 is very small. Similarly, the optical variance reduces with azimuth angle – waves are most easily seen by looking from their front or back than from a side look.

Numerical integration is tedious but the equations are highly nonlinear and hard to simplify analytically. Instead an empirical approximation was found to be

$$\Gamma_F(\tau,\alpha,k) = \kappa(k)\left[\sin^2(\tau)\cos^2(\alpha)+\cos^2(\tau)\right]\left(1+\cos(2\alpha)\right)\left.\frac{\partial R}{\partial \gamma}\right|_{\tau} \qquad (10)$$

where the gradient of R is found analytically and κ(k) describes the wavenumber dependence

$$\kappa(k) = \frac{k^2}{2}\left[0.631+0.345*\exp\left(10.203(k-0.178)\right)\right]. \qquad (11)$$

The approximation was found to be good to within 30% for 65% of the full azimuth and tilt space (-π to π in azimuth and 0 to π in tilt) and within a factor or two for

80%, mimicking well the 9 orders of magnitude variations in expected MTF. Errors were focused on extreme viewing angles such as near along-crest and nadir views where the assumption of Fresnel-dominated viewing physics is not valid.

The theory above is a simplification based only on the process of Fresnel reflection of homogeneous incident radiance. Nevertheless, several points are apparent. Wave contrast is better when looking in the direction of the waves and near horizontal – looking directly down yields much reduced wave contrast. Also, the k^2 dependence due to the wave steepness dependent physics implies that shorter waves (larger k) dominate signals for the same amplitude, consistent with the observation that short chop 'clutters' optical wave measurements.

An attempt was made to test the expected MTF relationships using three-minute video collections of a 50 by 50 m region outside the surf zone. Incoming waves were normally-incident swell with a period of approximately 8 s. Data were collected for three different tilts while looking into the approaching waves ($\alpha = 0$), then at four relative azimuths from 0 to 90° relative to the wave approach and with a tilt of 70°. Figure 8 shows the square root of intensity variance band-passed between wave periods of 3 and 18 s versus viewing angles.

The results in Figure 8 show that wave contrast clearly falls off as tilt goes toward nadir ($\tau = 0$), in a way that is consistent with the square root of the MTF (which expresses variance ratios, not shown). However, no significant azimuth dependence was found.

3.4 Returned Argus Products

Once the geometry has been solved for each frame, images can be sampled as they would be for a fixed Argus camera [*Holman and Stanley*, 2007]. Image products such as time exposures are designed as a matrix of world locations specified by minimum and maximum x and y locations, spatial resolutions, Δx and Δy, and an assumed vertical level, usually mean sea level. For every [x, y, z] location in this matrix, the corresponding image coordinates, [U, V], are found using equation (1) and the particular frame geometry, then intensities are found by interpolation into the image. From the resulting 3D data cube I(x, y, t), time exposures are found by simple averaging over time, but brightest, darkest and variance images are also computed using the appropriate statistics. Figure 9 shows an example time exposure. Individual pier pilings would be smeared and not visible if correct geometries had not been found for each frame.

Different data collection types have different accuracy requirements for image geometries. For instance, time exposures are smeared by definition so can tolerate geometry error. On the other hand, longshore current estimation tracks smaller foam patches so needs more accurate geometries. To estimate longshore currents we collect time series data from all of the contiguous pixels in longshore transects. Figure 10 shows a set of such time stacks collected at five different cross-shore locations from x = 125 m at the bottom to x = 225 at the top. The slope of the foam streaks, i.e. the rate at which foam drifts in time, corresponds to the alongshore current and can be estimated objectively [*Chickadel et al.*, 2003]. In figure 10 we see currents around y = 600 going to the south at offshore locations (foam drifting to smaller y) and to the north near the beach.

Even with the residual error from poorly constrained geometries, for example, for 10:45 EDT on 10/08/15 (see Table 3), extracted time series can be quite useful. Figure 11 shows a runup time series taken from a cross-shore continuous stack of pixels at y = 600. While the variability of geometry is made apparent by the wander of fixed features by around 1 m, seen on the left, the wave runup looks quite reasonable.

Finally, we wish to determine whether UAVs can be used with the cBathy algorithm [*Holman et al.*, 2013], a method of estimating bathymetry from wave celerity. cBathy was designed for the fixed viewing angles of Argus installations and measures both hourly and, given a series of hourly collects, in a running average way using a Kalman filter that compensates for time-varying sources of noise. Kalman-filtered cBathy results compare very well with ground truth surveys, with an observed 0.19 bias and 0.51 m RMSE over an 800 by 1000 area based on 16 surveys over two years [*Holman et al.*, 2013].

Figure 12 shows an example cBathy result from the UAV collections (middle) and the corresponding hourly result (without Kalman filtering) from the Argus data collection closest in time (left). For comparison, a partial CRAB survey from the following day (10/09/15) is in included in the right panel. Depth estimates with estimated errors greater than 1.0 m have been set to zero (brown color). Both remote sensing results are very similar, with an outer sand bar at ~280 m and with replicated alongshore variability. The mean deviation between the UAV and Argus results was -0.15 m, the standard deviation difference was 0.51 m and the 95th percentile of the absolute error was 0.95 m. Interestingly, the bad region in the Argus result (brown in left panel) at ~x = 230 corresponds to the onset of breaking over the inner bar, a region that has been hard to sample with Argus but is fixed by Kalman filtering from other tide stages when this was not occurring, but was not a problem for the UAV data, perhaps an advantage for lower tilt angles. Both methods had some trouble with the bar crest area around x = 250 m where breaking begins and cBathy is known to have high error (and correspondingly high estimated error).

4.0 Discussion

The cBathy comparison in the preceding section was of limited scope but suggests that UAVs can produce results that are comparable to a fixed Argus station and may even have some advantages in measuring at the onset of wave breaking. But the cases studied were limited and we have intentionally not included quantitative comparison against actual surveyed data. As described in the original cBathy paper [*Holman et al.*, 2013], while Kalman filtered results compare very well with surveys, hourly estimates can have varied quality particularly depending on the amount of breaking (automatically handled using the Kalman filter). Thus it was felt that the best test would be to see if how UAV results compared to simultaneous Argus collections. Clearly more work needs to be done on both of these points.

The Fresnel-based modulation transfer function theory matches limited observations moderately well for tilt variations but not for azimuth. This subject also needs more work. We were surprised at how much spatial variability in optical variance there was through the sampled region that seemed to be caused by other reasons – what we thought might be a bland area that could be sampled from many angles actually had a great deal of structure. Further tests should be run at greater distance from the surf zone and pier (hence well out over the water, a bit of a danger when flying UAVs). While the Fresnel reflectivity model might be reasonable for low grazing angles (near horizontal) that are used by Argus, it is apparent that at steeper viewing angles there are other sources of radiance that help us see waves. This also needs further work.

During this experiment, data were also collected using a Phantom 2 Vision Plus (P2VP), the immediate predecessor of the Phantom 3. Differences became apparent in terms of imagery, stability and metadata. The exif metadata stored in snapshots by the P2VP only records position to the nearest second of latitude and longitude, an approximately 30 m resolution in latitude. This error is too large to be useful. Camera pointing stability has also been improved in the Phantom 3 over the

Phantom 2. Station-keeping for the Phantom 3 is more accurate by factor of four while the viewing angles are generally more stable by roughly a factor of two. In addition, the Phantom 2 usually exhibited roughly one-minute oscillations in all geometry variables, presumably associated with tuning details in the stabilizing feedback loops (still, they are quite good). Finally, video image quality for the P2VP systems is good at 1080 pixel width but the P3P is much better at almost 4000 pixels across. The P2VP also has a great deal of barrel distortion that must be compensated for while the P3P sensor is much less distorted (see the horizon in Figure 2).

Many of the above results can help determine the best conops for data collection using a quadcopter for nearshore measurements when GCPs are limited in count and distribution across the image. The P3P has excellent autopilot positioning and can be considered fixed in space, even in strong winds. The position reported by the exif metadata is typically within 5 m of the true position but has a wander in pointing angles of several tenths of a degree (standard deviation). This residual wander can be corrected by observing at least one GCP with at least two being preferred. The resulting photogrammetry will be good in the vicinity of the GCPs but errors can grow away from regions of geometric control. Few natural beach locations will allow two dimensional placement of control points, so an assumption of either camera position from metadata or camera roll being some fixed value (hopefully zero) will make the solution possible. It is recommended that the camera roll be adjusted to zero, if possible, for this purpose.

Wave contrast decreases as the camera tilt is reduced toward nadir so that the signal becomes dominated from upwelling radiance from the water column rather than the surface. There is no point in aiming the camera down since you will both lose wave contrast and also area of coverage. The anticipated loss of signal with relative azimuth was not observed in data – waves continued to be easily visible even looking along-crest. Thus, along-coast camera look angles may allow extensive field of views as well as options for control points at intermediate and long distances.

5.0 Conclusions

We have investigated the potential for a low-cost quadcopter to collect image products and derived measurements of the nearshore that are comparable to Argus fixed cameras. In contrast to the increasing use of UAVs for land-based topographic measurements, nearshore applications will always have limited and poorly distributed ground control. In addition, in order to obtain the needed dwell, videos are often shot from a fixed location so have no motion to allow the use of popular structure from motion turn-key software. Full geometry solutions with just four, non-collinear control points are repeatable to 0.21 m. The UAV autopilot was found to have excellent station-keeping characteristic with a standard deviation in horizontal and vertical position of ~0.20 and 0.53 m, respectively, and in viewing angles about 0.25° with azimuth control being 0.38° standard deviation. UAV metadata on GPS position was found to be accurate to 5 m. A set of procedures are described and tested that substitute meta data or an assumption about camera roll to allow photogrammetric solutions with as few as two, or even one, control point with a typical horizontal positioning accuracy of imaged features of 10 m given a reasonable choice of control point location. A model for the dependence of wave contrast on viewing angles is discussed and found to be reasonable for mid to high tilt angles (toward the horizon) but not for the azimuthal dependence of viewing angle relative to the direction of wave propagation. Image products and quantitative measurements such as bathymetry and longshore currents can be found.

6.0 Acknowledgments

We would like to thank the staff at the USACE Field Research Facility for tremendous field support and Richie Slocum for insight on image rectification from the UAV. The BathyDuck experiment was supported by the Assistant Secretary of the Army for Acquisition, Logistics and Technology (ASAALT) Military Engineering RDT&E 6.2 and 6.3 Austere Entry Programs and by the USACE Coastal Field Data Collection Program. RAH was funded by the Office of Naval Research, grant number N00014-10-1–0932. Thanks as always to John Stanley for keeping everything running.

7.0 References

Birkemeier, W. A., and C. Mason (1984), The CRAB: A unique nearshore surveying vehicle, *Journal of Survey Engineering*, *110*, 1-7.

Brouwer, R. L., M. A. De Schipper, P. F. Rynne, F. J. Graham, A. J. H. M. Reniers, and J. H. MacMahan (2014), Surf zone monitoring using rotary wing unmanned aerial vehicles, *Journal of Atmospheric and Oceanic Technology*, *32*, 855 - 863, doi:10.1175/JTECH-D-14-00122.1.

Chickadel, C. C., R. A. Holman, and M. F. Freilich (2003), An optical technique for the measurement of longshore currents, *Journal of Geophysical Research*, *108*(C11), 3364.

Hartley, R., and A. Zisserman (2003), *Multiple view geometry in computer vision*, second ed., 665 pp., Cambridge University Press.

Holman, R. A., and M. C. Haller (2012), Remote sensing of the nearshore, *Annual Reviews of Marine Science*, *5*, 95-113, doi:10.1146/annurev-marine-121211-172408.

Holman, R. A., K. T. Holland, D. M. Lalejini, and S. D. Spansel (2011), Surf zone characterization from UAV imagery, *Ocean Dynamics*, doi:10.1007/s10236-011-0447-y.

Holman, R. A., N. G. Plant, and K. T. Holland (2013), cBathy: A robust algorithm for estimating nearshore bathymetry, *Journal of Geophysical Research*, *118*, 1-15, doi:10.1002/jgrc.20199.

Holman, R. A., and J. Stanley (2007), The history and technical capabilities of Argus, *Coastal Engineering*, *54*, 477-491.

Pennucci, G., D. C. Conley, and R. Holman (2007), Rapid environmental assessment from small unmanned aerial vehicles, paper presented at 22nd International Conference on Unmanned Aerial Vehicle Systems, University of Bristol, Bristol, England, 2007.

Taborda, R., and A. Silva (2012), COSMOS: A lightweight coastal monitoring system, *Computers and Geosciences*, *49*, 248-255, doi:10.1016/j.cageo.2012.07.013.

Toth, C., G. Jozkow, and D. Grejner-Brzezinska (2015), Mapping with small UAS: a point cloud accuracy assessment, *Journal of Applied Geodesy*, *9*(4), 213-226, doi:10.1515/jag-2015-0017.

Turner, I. L., M. D. Harley, and C. D. Drummond (2016), UAVs for coastal surveying, *Coastal Engineering*, *114*, 19-24, doi:10.1016/j.coastaleng.2016.03.011.

Walker, R. E. (1994), *Marine light field statistics*, 675 pp., John Wiley and Sons, Inc., New York.

Wolf, P. R. (1983), *Elements of Photogrammetry*, McGraw-Hill, New York.

Date	Time	Δ_x (m)	Δ_y (m)	Δ_z(m)	σ_x (m)	σ_y (m)	σ_z (m)
10/08/15	11:54	-11.11	5.15	10.52	0.34	0.27	0.13
10/07/15	15:29	-1.49	-7.68	12.22	0.16	0.13	0.28
10/07/15	15:29	-1.80	-7.84	12.22	0.16	0.16	0.30
10/08/15	11:58	-7.38	-0.10	9.37	0.12	0.21	0.19
10/08/15	12:01	-1.38	-5.36	8.85	0.27	0.07	0.34
MEAN		-4.63	-3.17	10.64	0.21	0.17	0.25

Table I. Errors associated with GPS camera location metadata as judged by solutions for camera location using four digitized GCPS for each of the five images studied. Δ and σ values are the mean and standard deviation of differences computed over the three replicates.

Knowns	Unknowns	Number of GCPs	Mean error (m)	Median error (m)	95% error (m)	Max error (m)
None	x_c, y_c, z_c az, tilt, roll	4	0.9	0.6	1.7	4.9
Roll	x_c, y_c, z_c az, tilt	3	1.8	0.6	10.4	23.9
x_c, y_c, z_c	az, tilt, roll	3	8.5	7.2	20.0	25.3
		2	10.2	6.9	30.6	51.8
x_c, y_c, z_c roll	az, tilt	3	7.1	6.9	13.1	15.8
		2	7.6	6.2	18.1	33.7
		1	10.0	5.5	39.0	61.0
x_c, y_c, roll	z_c az, tilt	3	10.1	4.4	41.5	81.6
		2	11.0	5.4	44.7	138.0

Table 2. Ground distance error statistics based on five test images for different choices of known and unknown variables.

Date	Time (EDT)	Dwell (s)	σ_x (m)	σ_y (m)	σ_z (m)	$\sigma_{azimuth}$ (deg)	σ_{tilt} (deg)	σ_{roll} (deg)
10/01/15	1529	791	0.21	0.17	0.36	0.74	0.37	0.39
10/07/15	1530	560	0.21	0.31	0.27	0.48	0.18	0.30
10/08/15	0936	960	0.29	0.36	1.47	0.47	0.39	0.68
10/08/15	1014	182	0.10	0.16	0.12	0.13	0.09	0.18
10/08/15	1018	181	NA	NA	0.45	0.28	0.14	NA
10/08/15	1022	181	NA	NA	0.50	0.51	0.10	NA
10/08/15	1045	182	NA	NA	0.94	0.44	0.33	NA
10/08/15	1049	181	0.15	0.26	NA	0.40	0.19	0.20
10/08/15	1053	181	0.06	0.21	NA	0.17	0.21	0.06
10/08/15	1057	182	0.14	0.21	0.12	0.27	0.06	0.07
Means			0.17	0.24	0.53	0.38	0.20	0.26

Table 3. Statistics of variability expressed as standard deviations (σ) for the six geometry variables for ten data runs of varying dwell. NA values correspond to variables that were fixed during the geometry solution as discussed in section 3.1. The third run was significantly noisier than others for no known reason.

Figure 1. A Phantom 3 Professional quadcopter. The camera is integrated into the unit using a specialized stabilized gimbal.

Figure 2. Example snapshot used in testing the accuracy of GPS metadata. The four selected GCP locations are shown by black circles.

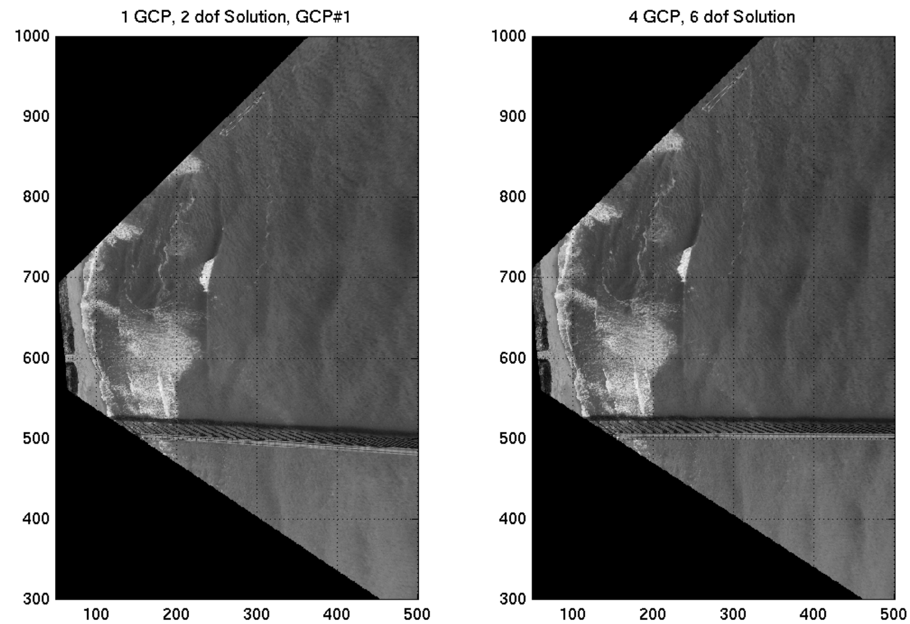

Figure 3. Rectifications of the first test image. The left image has used a 2 DOF geometry solution found with only one control point (not a least squares solution) while the geometry of the right image was solved using all 6 DOFs using four control points, and is accurate to about 1 m. The orientation of the pier and location of the CRAB (elongated triangle near x = 275, y = 900) exemplify the errors in the lower accuracy, left rectification.

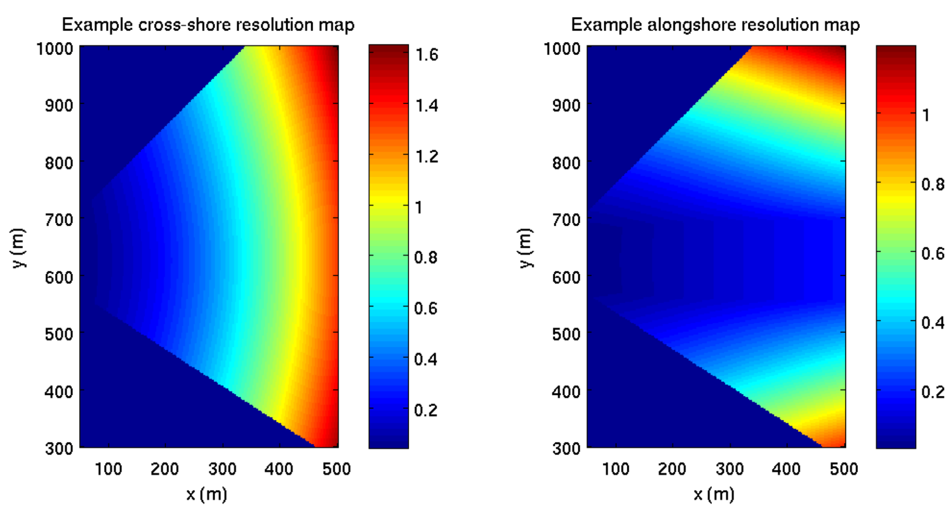

Figure 4. Resolution maps for the snapshot shown in Figure 3. Left panel shows the range resolution (away from camera) while the right panel shows the cross-range component. Both are expressed in meters.

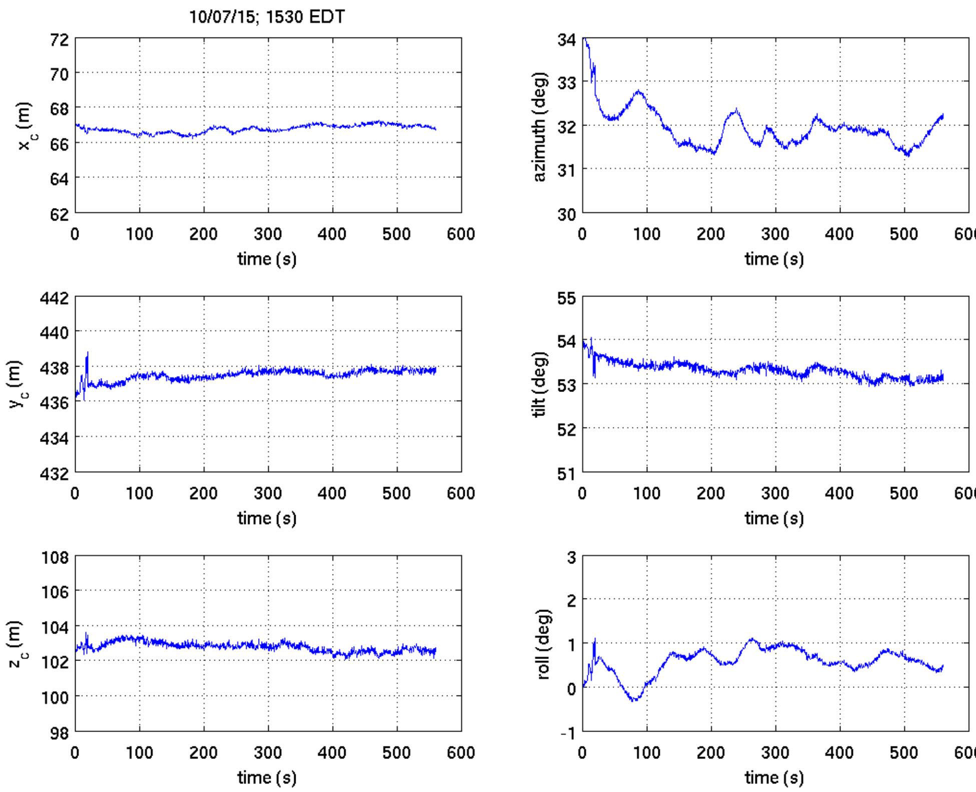

Figure 5. Stability of camera position (left panels) and viewing angles (right panels) for an example 560 second data run at 1530 EDT, 10/07/15.

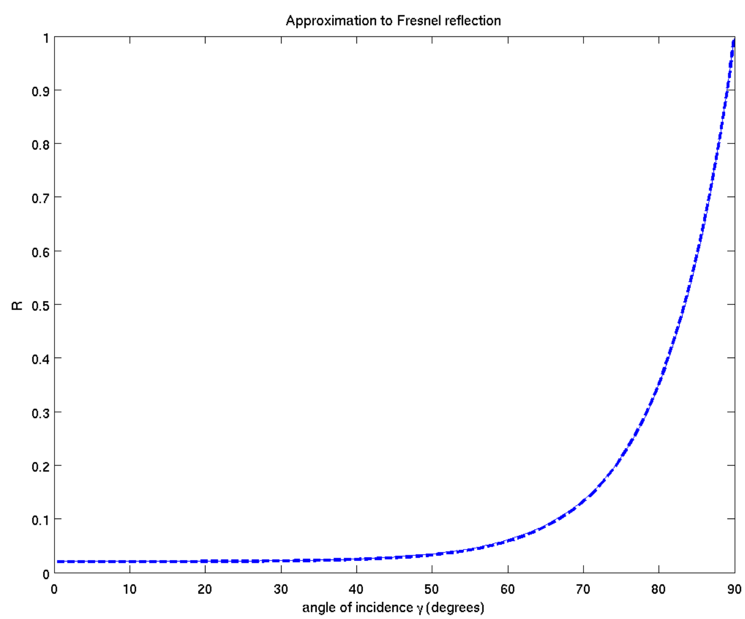

Figure 6. Dependence of reflection coefficient, R, on the angle of incidence, γ. The thin solid line is the full Fresnel equation while the (indistinguishable) thicker dashed line is the approximation in equation (5).

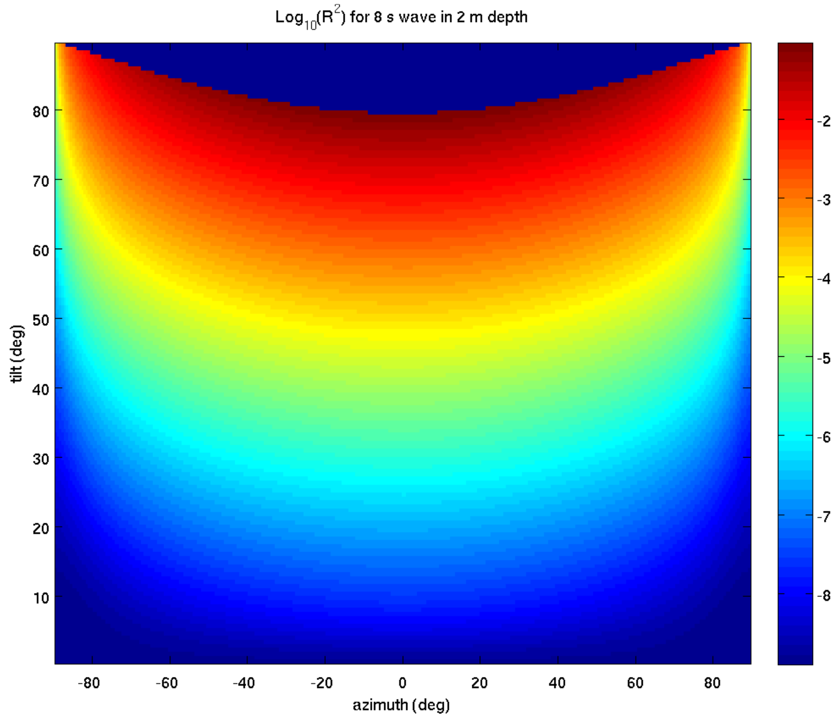

Figure 7. Squared reflectance for a test case of 8 s waves in 2 m depth, shown as $\log_{10}$ of the computed values. The dark area at the top is due to part of the waves being shadowed for near horizontal look angles.

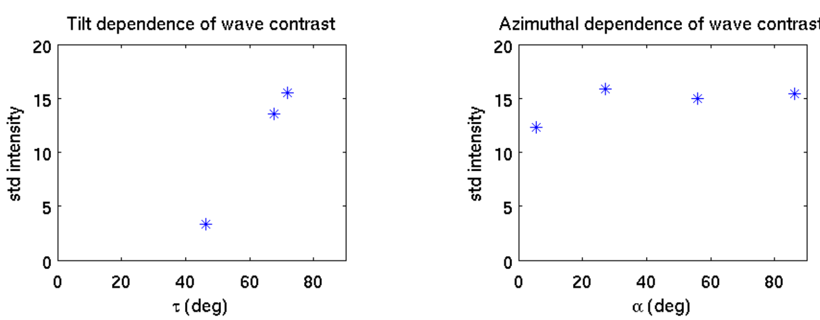

Figure 8. Standard deviation of band-passed intensity versus viewing angles.

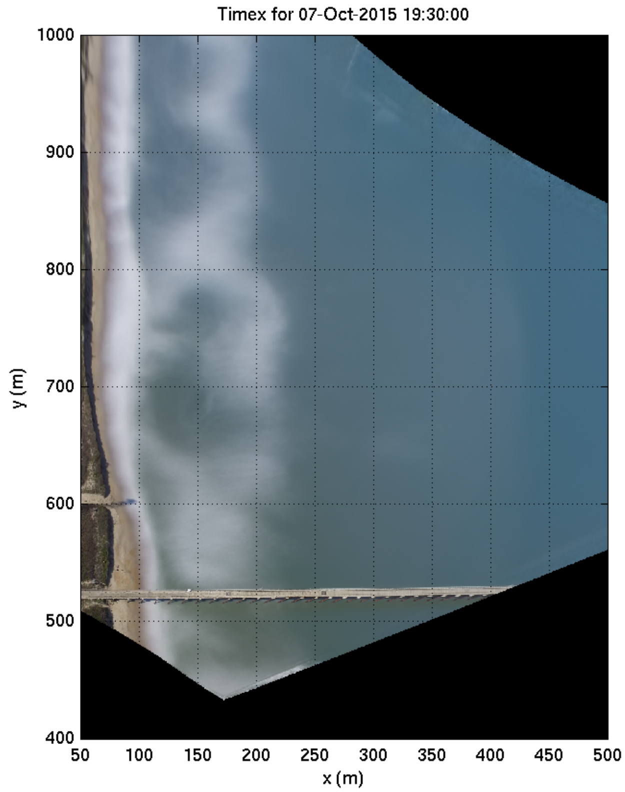

Figure 9. Example ten-minute time exposure showing the pier (horizontal feature at y ~ 515 m) and the sand bar and shoreline morphologies for 10/07/15. The camera was located at the bottom left. The top of the field of view appears curved due to lens distortion.

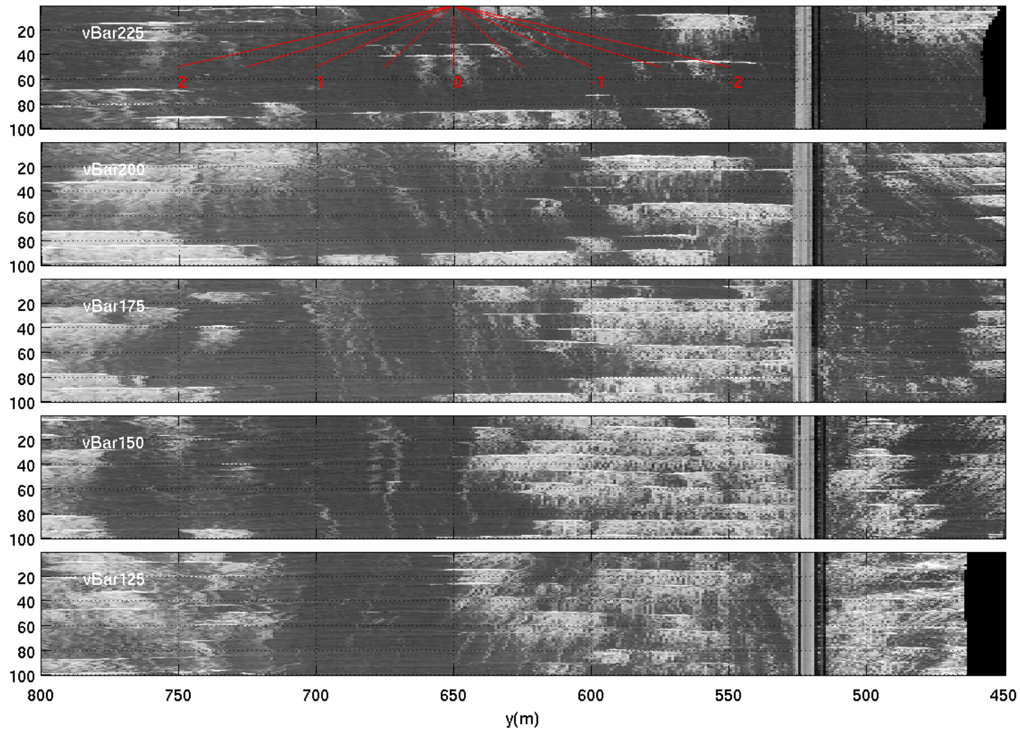

Figure 10. Time-space images from five alongshore lines from 125 to 225 m in FRF x coordinates (see Figure 9). For each panel, time increases down the page (marked in seconds). The pier is visible at y ~ 520. White indicates wave breaking and the thin streaks indicate residual foam. Sloping trajectories of thin foam streaks indicate advection of foam in the longshore direction. Slopes can be used to estimate those currents. Red lines in the top panel indicate the slopes for different currents.

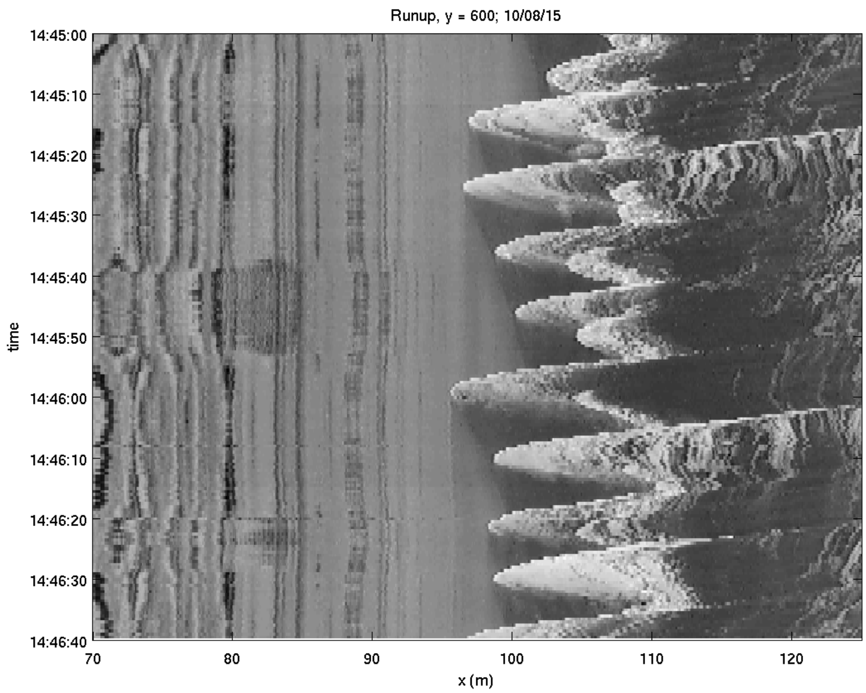

Figure 11. Runup time stack for y = 600 for 10/08/10:45 EDT, a run with poor geometry control. Wander in the geometry solution is evident on the left, as fixed features on the beach move somewhat. Yet the runup time series looks quite useable.

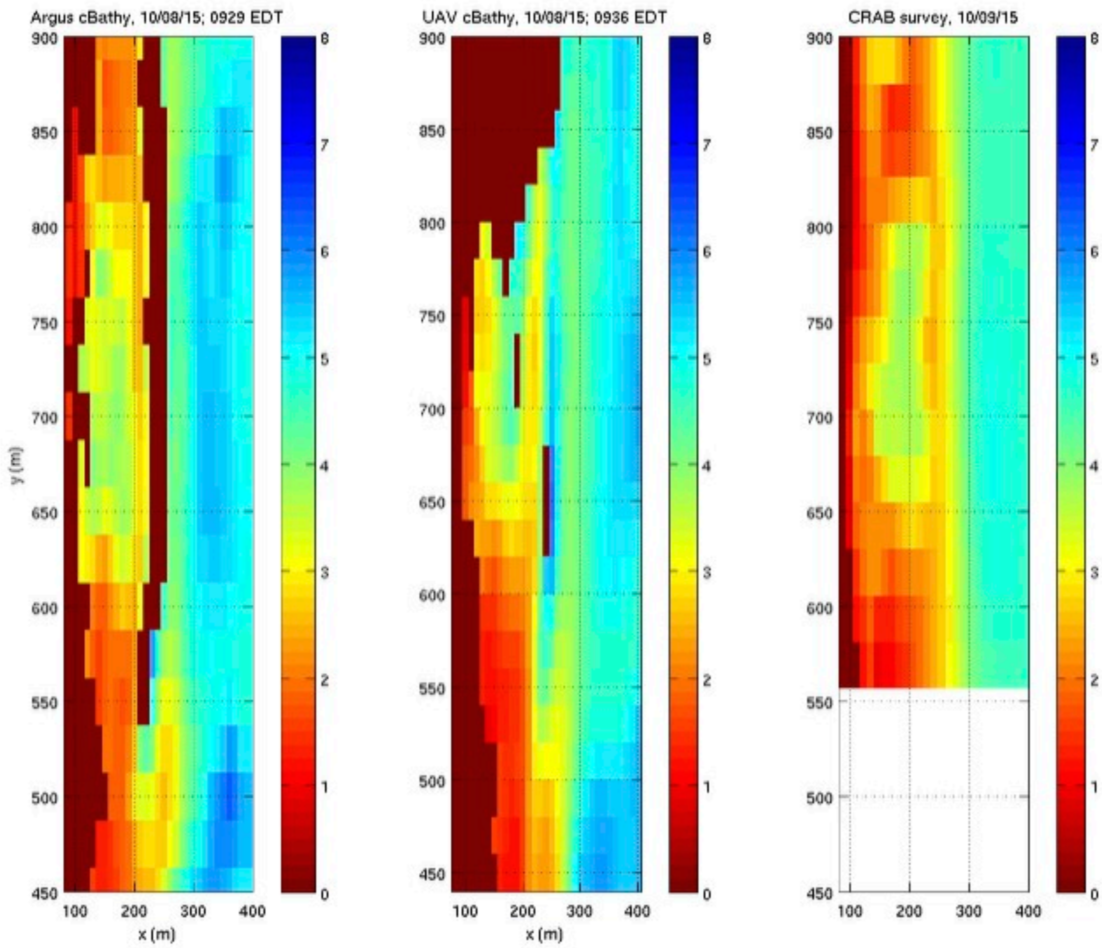

Figure 12. Comparison of example cBathy depth (m) results found using UAV data (middle) with that found using the fixed Argus camera (left). Values with estimated errors greater than 1.0 m have been omitted (shown as brown). For reference, a partial CRAB survey from the next day is show.