

An ROV Stereovision System for Ship Hull Inspection

Shahriar Negahdaripour, *Senior Member, IEEE* and Pezhman Firoozfam

Abstract—Ship hulls, as well as bridges, port dock pilings, dams, and various underwater structures need to be inspected for periodic maintenance. Additionally, there is a critical need to provide protection against sabotage activities, and to establish effective countermeasure against illegal smuggling activities. Unmanned underwater vehicles are suitable platforms for the development of automated inspection systems, but require integration with appropriate sensor technologies.

This paper describes a vision system for automated ship inspection, based on computing the necessary information for positioning, navigation and mapping of the hull from stereo images. Binocular cues are critical in resolving a number of complex visual artifacts that hamper monocular vision in shallow water conditions. Furthermore, they simplify the estimation of vehicle pose and motion which is fundamental for successful automatic operation. The system has been implemented on a commercial ROV, and tested in pool and dock tests. Results from various trials are presented to demonstrate the system capabilities.

Index Terms—Underwater Visual Inspection, stereovision, ROVs, and AUVs.

I. INTRODUCTION

UN-DERWATER inspection is an essential task for general maintenance and damage assessment of underwater structures. For example, inspection of ship hulls is necessary as part of periodic maintenance operations. This has become extremely critical with the threat that ships entering ports and harbors for commerce may serve as carriers of nuclear weapons, explosives, deadly chemicals and other hazardous materials, with mass destruction in highly populated cities, national landmarks, and other drastic damages at the nation scale as potentially target activities [50]. To combat this threat, deployment of existing technologies and the development of new ones are sought to implement search and detection systems that can provide no less than 100% success rate. Unlike regular hull maintenance that may be carried out by trained divers, inspection and search for hazardous and (or) deadly materials have to be done with submersible robotics platforms to avoid risk of human lives. In general, it is expected that the deployment of such vehicles, when highly automated, can provide a more effective and efficient solution.

Though AUV proponents may view them as “the” appropriate platform for underwater search and mapping, ROVs offer distinct advantages in human-assisted or human-supervised hull search operations [42]: 1) Real-time video and data transmission to the operator station enables him/her to revise the

mission on the fly or take over the operation, where necessary; 2) Division of labor can be readily achieved by the automation of many low-level tasks (as in precise navigation and the construction of a composite map), allowing the operator to concentrate on high-level critical components of the operation (e.g. target and object recognition, threat assessment, etc.).

In recent years, automated video-based servo, survey and mapping has become recognized as important capabilities in ROV deployment for seafloor and benthic habitat studies. This has led to worldwide effort on the development and implementation of visual servo technologies for implementation on various research ROVs, with and without integration with other sensory information. In addition to our work at the University of Miami, we are aware of the work at these other institutions¹: Heriot-Watt University (UK) [29], [40], [46], Instituto Superior Tecnico (ISR, Portugal) [17]–[19], MBARI-Stanford group [10], [11], [28], [30], IFREMER/INRIA [26], [47], Australian National University (RSL) [45], [48], Univ. of Sydney (ACFR, Australia) [31], University of Tokyo/Sigapore [2], [8], Universitat de Girona [15], [16], Universitat de les Illes Balears [39].

The earlier work addressed the use of images for optical station keeping of ROVs (e.g., [27], [29], [34], [35], [37]), while more recent work has been aimed at the realization of image mosaicing capability to map the sea floor and various benthic structures for fisheries, reef studies, archeology and various other marine science applications (e.g., [7], [10], [11], [16], [18], [38], [43], [49]). Some research has been aimed at the use of mosaics as visual maps for improved positioning and local navigation (e.g., [17], [19], [36], [38]). Integration with other sensors has been explored to improve the positioning and photo-mosaicing capabilities; e.g., [3], [31], [41], [47]).

In deep sea, shading flows induced by the motion of artificial sources installed on the vehicle have to be overcome in processing of images for video servo and image mosaicing. In shallow waters, near the sea surface sources of disturbance are of primary concern. First, surface waves often cast complex moving shadow patterns on the targets that are being imaged/inspected. Other difficulties arise from the movement of floating suspended particles and water bubbles that are commonly present in most coastal and harbor waters. On top of these, constant disturbances from wave actions contribute to continuous complex vehicle motions – involving all six degrees of freedom – that are often difficult to estimate from monocular cues; even without the low quality and contrast of

The authors are with the Electrical and Computer Engineering Department, University of Miami, Coral Gables, FL 33146 USA (e-mail: shahriar@miami.edu; pezhman@miami.edu).

¹We have cited only selected publications that describe sample activities at these research centers. Readers are strongly encouraged to refer to the individual web sites for detailed information.

underwater imagery and the complex shadow motions. Where the vehicle has to fight steady currents, reduced thruster power for control action takes away from the vehicle maneuvering capability. As a consequence of these various factors, the vision system has to process images from a very complex motion video. Adding insult to injury, poor control performance would typically lead to a longer operation than necessary. This offers the vision system the unfortunate opportunity to spit out more erroneous motion estimates that, once integrated over a longer time, produce even larger position drift errors.

Because of these factors, the use of existing monocular vision-based technologies –often suitable for deep sea tasks –can be effectively ruled out because of the unfriendly conditions in most shallow harbor waters. Interestingly, the use of binocular cues offers unique potentials: As we will elaborate, it allows us to take advantage of visual cues in such a way that the above complexities can be made much less effective in destabilizing the system performance. This motivated us to explore the use of a stereovision system for ship inspection, an operation carried out by ROVs in the shallow waters of ports and harbors. The investigation has led to the development of a vision system that provides three key capabilities in the automatic inspection of ship hulls and other underwater structures: Positioning and navigation, 2-D mapping with a photo-mosaic, and the 3-D reconstruction of the target structure. As always, the binocular cues give an instantaneous 3-D map of the target, thus enabling the detection of foreign objects attached to the subsea structure. Automatic navigation is achieved by estimating the 6 degrees of freedom (d.o.f) in ROV position directly from visual information. Binocular cues also give the instantaneous measurements of the ROV distance and orientation with respect to the structure, which can be used for positioning and trajectory control. Finally, information about the ROV motion, determined from consecutive frames of the video, is used for image alignment in constructing a photo-mosaic.

Human-assisted or supervised search and inspection are significantly facilitated by utilizing a 2-D mosaic and a 3-D structural map of the imaged object displayed on the operator monitor, with the additional capabilities of automatic navigation and vision-based station keeping developed previously [37], [38]. The stereovision system has been implemented on a Phantom XTL² and the capabilities have been successfully demonstrated in pool and dock tests. Results from various real-time operations are presented to demonstrate the system performance.

II. TARGET-BASED NAVIGATION AND VISUAL SERVO

In automated navigation, the ROV is generally required to execute a desired trajectory under computer control based on the information from various positioning sensors. In a common scenario in ship inspection, the vehicle may be programmed to navigate along the side, while maintaining a frontal view and a fixed safe distance D_s relative to the hull. As the hull surface may bend, the vehicle would turn with it. This suggests that it is

favorable to seek pose and distance measurements relative to the ship, rather than absolute positions and orientations which may be harder to determine or more costly with expensive sensors.

In particular, some positioning sensors commonly deployed for absolute measurements, such as INS and acoustic-based devices, have certain limitations: drift error, acoustic clutter, risk of sabotage when deploying permanent fixed bottom transponders in security applications, complexities in the calibration of temporary transponders suspended from the ship, etc. An optical imaging system, deployed for target visualization and mapping, may simultaneously provide the necessary information for automatic navigation. In particular, video-based servo and mapping has been two popular applications of vision techniques in underwater, and many systems have been developed for the mapping of benthic habitats. The fundamental problem is to estimate the frame-to-frame motions from the corresponding pair of images.

We next discuss some significant challenges that are to be overcome by a vision system in shallow water operations. These support the conclusion that existing techniques utilizing visual cues from a single camera (monocular vision) are practically ineffective for automated ship inspection. We then explain why the use of a stereovision system can significantly reduce the complexities.

A. Complexities of a Monocular Vision System in Shallow Water Conditions

Moving (cast) shadows and non-uniform shading patterns that vary temporally are responsible for some major complexities in the processing of video imagery. In some cases, e.g., where the target surfaces have weak texture, camera motion-induced visual cues can become dominated by the image shading artifacts. It therefore becomes difficult to decouple the image motion induced by the camera motion from the shading flow.

In deep sea operations, absence of natural lighting requiring the use of artificial sources with finite power is mainly responsible for the challenge of having to overcome time-varying shadow artifacts in processing of a motion video. However, to the extent that the frame-to-frame platform movements are smooth and slow, so are the temporal variations in shading over the image. Consequently, methods can be devised, tuned to the properties of the illumination source, to filter out the lighting effects [14]. In operations near the sea surface, floating bubbles and suspended particles, nonuniform illumination from natural lighting, cast shadows from surface waves and reflections from the target surfaces are factors that lead to rather complex visual motion artifacts that can dominate the image variations due to the motion of the camera/vehicle. In loose terms, deep sea image shading artifacts can be viewed as deterministic and predictable, while near surface events are unpredictable and somewhat stochastic. Furthermore, the higher turbidity of most harbors, in comparison to deep sea conditions, result in images with much less contrast and poorer quality. These various factors collectively make it rather difficult to match points or regions in consecutive video frames –a pre-requisite step

²Trademark of Deep Ocean Engineering, San Leandro, CA

for computing the vehicle motion and (or) registering these frames (e.g., to construct a photo-mosaic). The problems are compounded by the fact that the surface wave and current actions can move the vehicle with all 6 degrees of freedom. It is practically impossible to estimate the 6 d.o.f motions with acceptable accuracy from unreliably matched features. Two more factors work against us. The effective f.o.v is reduced in underwater, the target surfaces are usually (locally) flat, both of which are extremely undesirable for discerning the motion components due to the well-known translation-motion ambiguity [1].

Now, video enhancement and restoration methods may be devised to process the entire data from an operation in order to discriminate between the signal component and the disturbance artifacts. However, this is impractical for real-time operations.

Estimation of the various degrees of freedom (d.o.f.) in the motion of the vehicle is important for both position control and target mapping. As stated, we seek the instantaneous pose (position and orientation) of the ROV relative to the target, rather than its absolute 3-D position, as the most relevant information.

Most small-size ROVs, including our Phantom XTL, have four—two aft, one lateral and one vertical—thrusters for XYZ translations and heading change. The video camera is commonly installed in front, and can be pointed anywhere from the forward to downward directions. The *ideal mode for visual servo*, when the vehicle navigates along the seafloor, is the down-look configuration; see fig. 1a. The main reason is that the 4 d.o.f. in the vehicle motion, controllable through the proper signals to the 4 thrusters, are exactly the same 4 d.o.f. that can be estimated most reliably from the video frames (of the seafloor). One can claim that the controllable system states are all observable. How about the uncontrollable states, namely the pitch and roll motions? While these can theoretically be determined also from video, the estimation is seldom robust and accurate, particularly where the target scene (seafloor) is relatively flat and is imaged over a relatively small field of view³ [1]; that is, topographical variations are small compared to the distance to the sea floor. Accordingly, the most ideal scenario to maintain positioning accuracy by visual servo is to navigate with no, or very little, pitch and roll motion. To observe (estimate) these other motion components, inexpensive angle sensors are often sufficient. In this case, the video can be rectified to correct for (stabilized w.r.t) pitch and roll motions, before processing to estimate the other 4 d.o.f., providing all the necessary information for positioning and mosaicing [3].

For ship inspection, the ROV would need to move along the sides, while maintaining a constant distance and orientation relative to the hull. In one scenario, an extra camera is installed in a side-look arrangement (see fig. 1b), while the vehicle moves forward/backward along the hull. Alternatively, the existing camera may be pointed in forward-look configuration, while the vehicle moves sideways (left and right) to map the ship. In either of these situations, the change in ROV

heading corresponds to the pan motion of the camera, which cannot be estimated with good accuracy from video when coupled with side-to-side translation. Unfortunately, the heading change cannot be reliably measured from typical compasses due to the magnetic masking. In most cases, the reliable estimation of the dominant motion components by the vision system is reduced to 2-3 degrees of freedom.

B. Stereovision Solution

A stereovision system can overcome most of the complexities of mapping and positioning with a monocular vision system, as we explain next. In the remainder, we assume that the relative positions of the stereo cameras has been determined by external calibration [51], and any inaccuracy in the parallel alignment is corrected by stereo rectification [13], [22]. Clearly, we still have to overcome the near surface disturbances, in addition to having to process images with a lower quality due to the turbidity factor. However, the complexities from cast shadows are of little concern, and in fact are exploited to our advantage in improving the positioning accuracy. Consider as an example a surface with very weak texture, which may be a poor candidate for temporal and spatial matching. Suppose the surface is projected with a relatively strong non-uniform shading pattern. This provides us with surface markings for stereo matching, just as in 3-D reconstruction from structured light patterns in industrial application; e.g., [4], [44], [52].

To explain more formally, consider a nonuniform moving pattern $W(X; Y; t)$ on the target surface, at some time t . It projects onto a local region $w^l(t)$ in the left image with intensity $I^l(x_w^l(t); y_w^l(t))$, where $f(x_w^l(t); y_w^l(t))g$ denotes the coordinates of pixels within the region $w^l(t)$. At the next sampling time $t + dt$, the same pattern projects onto region $w^l(t + dt)$ comprising pixels $f(x_w^l(t + dt); y_w^l(t + dt))g$:

$$x_w^l(t + dt) = x_w^l(t) + dx_w(t) \quad \text{and} \quad y_w^l(t + dt) = y_w^l(t) + dy_w(t);$$

where $dx_w(t); dy_w(t)g$ is the motion of pixel $f(x_w^l(t); y_w^l(t))g$. Suppose the pattern varies significantly in intensity due to some strong temporal effect. If the ROV motion is also complex with 6 d.o.f., the region shape would also change somewhat⁴. Then, it is the case that $I(x_w^l(t); y_w^l(t))$ and $I(x_w^l(t + dt); y_w^l(t + dt))$ have large enough differences, and $w^l(t)$ and $w^l(t + dt)$ would differ in shape somewhat due to projection deformations. It thus becomes difficult to match the two regions, based on either geometric or radiometric cues. The same region $W(X; Y; t)$ appears in the right image at some disparity, inversely proportional to the distance to the corresponding points on the target surface. Without loss of generality, assume that the surface is relatively flat at some desired distance D_s from the ROV. There should be minimal distortion between the two views $f(w^l(t); w^r(t))g$ of the region, since the surface is viewed (close to) frontally. Therefore, the corresponding region $w^r(t)$ in the right image consists of points $f(x_w^r(t); y_w^r(t))g$, where

$$x_w^r(t) \approx x_w^l(t) \quad \text{and} \quad y_w^r(t) = y_w^l(t);$$

³Recall that the effective f.o.v. is reduced in underwater imaging.

⁴This factor is somewhat immaterial with respect to the radiometric effects.

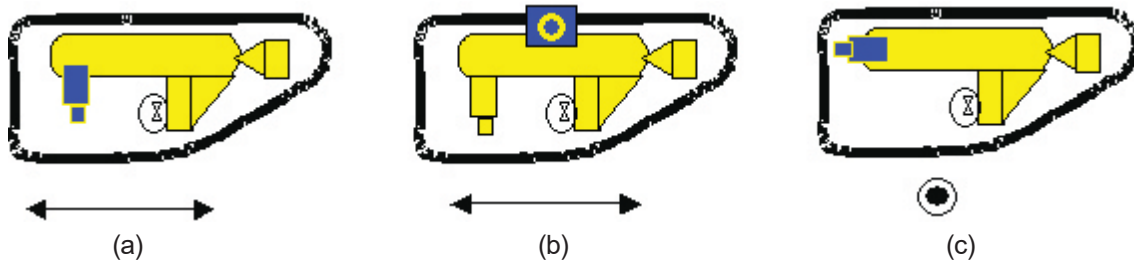


Fig. 1. Down-look camera configuration for seafloor mapping (a), and two potential arrangements for hull inspection (b-c) with arrows below giving the corresponding motion direction.

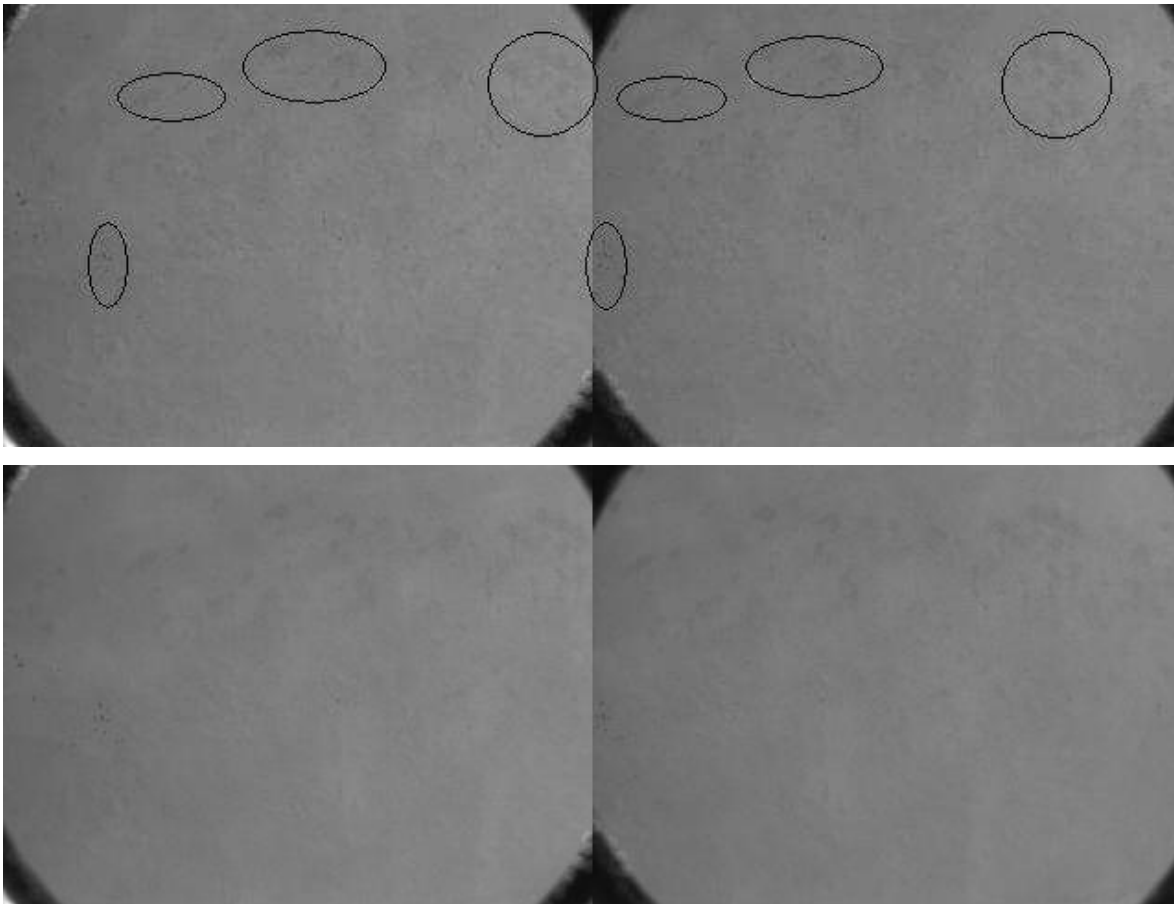


Fig. 2. Two consecutive stereo pairs of a relatively feature-less pool wall. Encircled regions can be readily matched in the first pair. Though they can also be matched in the second view, some have a slightly different visual appearance because of temporal shading variations. (see section 2.B for details).

where f is the effective focal length of the cameras (made to be the same, if different, through the a priori calibration process), and b is the stereo baseline. Furthermore, the two views should have more or less similar intensities:

$$I^l(x_w^l(t); y_w^l(t)) \approx I^r(x_w^r(t) + f b = D_s; y_w^r(t)):$$

The spatial matching problem, only a 1-D search with relatively smooth disparity variations, is significantly simpler than the temporal matching for the monocular case. With the stereovision system, we still have to perform temporal matching between consecutive ROV positions, to estimate the ROV motion. However, as explained in detail below, this motion problem is also simplified, since we apply the feature/region matching to rectified (frontal) views with relatively constant displacements over the image.

Fig. 2 depicts two consecutive stereo pairs of a very low-textured surface in shallow water (a swimming pool wall), taken during a system testing operation. Certain matching texture patterns have been encircled in the first stereo pair. The same regions can also be identified and matched in the second stereo pair, though some have (slightly) different visual appearances from one time to the next; thus, are (somewhat) harder to match temporally.

Another advantage of stereovision comes from the ability to estimate with much higher precision the 6 d.o.f. positioning information (relative to the ship) by utilizing the binocular cues, as described in the next section. More accurate sensory information enables better control in maintaining a more stable platform. Consequently, we would have to process a less complex motion video due to a smoother vehicle motion.

III. STEREOVISION-BASED POSITIONING

The ROV positioning, by determining the pose and (or) the frame-to-frame movements, is significantly simplified by decomposing it into two simpler problems to be solved consecutively: 1) Three d.o.f by stereo matching applied to the left and right stereo pairs $I^l(t)$ and $I^r(t)$ at time t ; 2) Three other d.o.f by frame-to-frame registration. Let the 3-D vector $(-x(t); -y(t); -z(t))$ define the angle that describes the orientation of the ROV at sample time t relative to the target surface. Similarly, $X(t)$ and $Y(t)$ denote the estimated horizontal and vertical positions with respect to the starting point, and $Z(t)$ is the instantaneous distance along the optical axis from the target surface. The 3-D vectors $(!_x(t); !_y(t); !_z(t))$ and $(T_x(t); T_y(t); T_z(t))$ denote the estimated frame-to-frame rotational and translational motions at time t .

Assume that we want to maintain a frontal view (heading $-y(t) = 0$) and a fixed safe perpendicular distance $Z(t) = D_s$ from the ship. Ideally, two other orientation angles, pitch $-x(t)$ and roll $-z(t)$, are to remain at zero. Ensuring these latter conditions requires a platform with pitch and roll control, while most small-size ROVs have a truster configuration that allows the control of only positions in 3-D and heading. (Pitch and roll stability is achieved primarily through the design of the platform, but cannot be controlled through thruster action.) In our system, we use these measurements not to control/correct the ROV orientation, but to do pitch and roll stabilization of the video that is processed for motion estimation, image registration and finally mosaic construction.

We determine at each instant three positioning components from binocular stereocues—namely from the disparity of $I^l(t)$ and $I^r(t)$. These comprise the relative heading $-y(t)$, pitch angle $-x(t)$, and the distance $Z(t)$ with respect to the local hull surface, being viewed. Using the measured angles and distance, we rectify the stereo images: We construct rectified left $I^l(t)$ or right $I^r(t)$ video⁵ that gives frontal views of the hull at the desired constant distance D_s ; see fig. 3 (top). We finally determine the translations in X and Y and roll motion $!_z$ from rectified views.

For simplicity, assume the target is a plane with equation $Z = q_0 + q_1X + q_2Y$ is the coordinate system of the left camera⁶. This can be written nn^TXX , where $nn^T = \frac{1}{q_0^2}(q_1; q_2; 1)$ is the surface normal vector and $PP = (X; Y; Z)$ is a point on the plane. As stated, we can readily determine the q_i 's (elements of nn^T) from the stereo disparities $d = x^l_j - x^r_j$ at points $pp^l = (x^l; y^l; f)$. The underlying constraint

$$d = f b = Z = f b (nn^T pp)$$

is linear with three unknowns. Disparities at a minimum of three points, which do not lie on a single image line, yield a unique solution⁷. We establish from q_1 and q_2 the orientation and from q_0 the distance of the target surface with respect to the camera.

⁵Only left or right video is sufficient

⁶This concept can be generalized to other locally smooth shapes since we can determine the target shape from stereo cues. Non-smooth surfaces can also be treated provided we establish a suitable definition for a frontal look.

⁷In practice, we use more points in a least-square formulation, or RANSAC-based implementation [12].

To construct the frontal view, we need to calculate for every pixel $pp = (x; y; f)$ in the original image its position $pp^0 = (x^0; y^0; f)$ in the rectified image; see fig. 3 (middle). There are an infinite number of transformations to accomplish this, as one degree of freedom – rotation $!_z$ about the viewing direction – can be chosen arbitrarily. Since we will be determining this motion in the next step, we choose the unique solution comprising no rotation about optical axis. In other words, the rotation is performed about the axis $rr = (r_1; r_2; 0)$ by angle μ :

$$r_1 = j \frac{q_2}{q_1^2 + q_2^2}; \quad r_2 = q \frac{q_1}{q_1^2 + q_2^2}; \quad \mu = \tan^{-1}(q_1^2 + q_2^2)$$

This maps the points $PP = (X; Y; Z)$ to $PP^0 = (X^0; Y^0; Z^0)$ on the frontal plane $Z^0 = q_0 \cos \mu$. For proper scaling, the rectified image points $pp^0 = (x^0; y^0)$ are constructed from $x^0 = j f X^0 = D_s$ and $y^0 = j f Y^0 = D_s$.

Having constructed the rectified view $I^l(t)$ at each ROV position, consecutive frames $I^l(t)$ and $I^l(t + dt)$ are processed to calculate the displacements T_x and T_y and roll motion $!_z$ about the camera optical axis. When the rotation is negligible, a simple 2-D correlation applied over the entire image provides some level of robustness with respect to small errors from the first step. Otherwise, we can deploy the 2-D motion model

$$pp^0(t + dt) = RR(!_z)pp^0(t) + TT_{xy}$$

This corresponds to an isotropic image transformation with image shift $TT_{xy} = f T_x; T_y g$ and rotation $!_z$. In fact, we can utilize a more general solution based on a similarity transformation by allowing a scale change s between the two views (e.g., if the frontal view rectification process does not perfectly yield images at the same distance from the target due to estimation error in the first step):

$$pp^0(t + dt) = sRR(!_z)pp^0(t) + TT_{xy}$$

The unknown motion components in either case can be computed from a number of linear methods, based on feature or regions matching, optical flow, or a direct method [37], [38].

A particular robust close-form solution utilizes a simplified 2-D form of the absolute orientation in photogrammetry [24]. By defining

$$cc^0(t) = (1=N) \sum_{i=1:N} pp^0(t);$$

we can determine the translation from $TT = sRR(!_z)cc^0(t);$

if the scale and rotation are known. The scale is readily determined from $s = \frac{\sum_{i=1:N} (pp^0(t + dt))^T pp^0(t + dt)}{\sum_{i=1:N} (pp^0(t))^T pp^0(t)}$,

and the planar rotation (angle $!_z$) can be computed from the equation

$$pp^0(t + dt) = RR(!_z)pp^0(t);$$

where

$$pp^0(t) = pp^0(t); \quad cc^0(t);$$

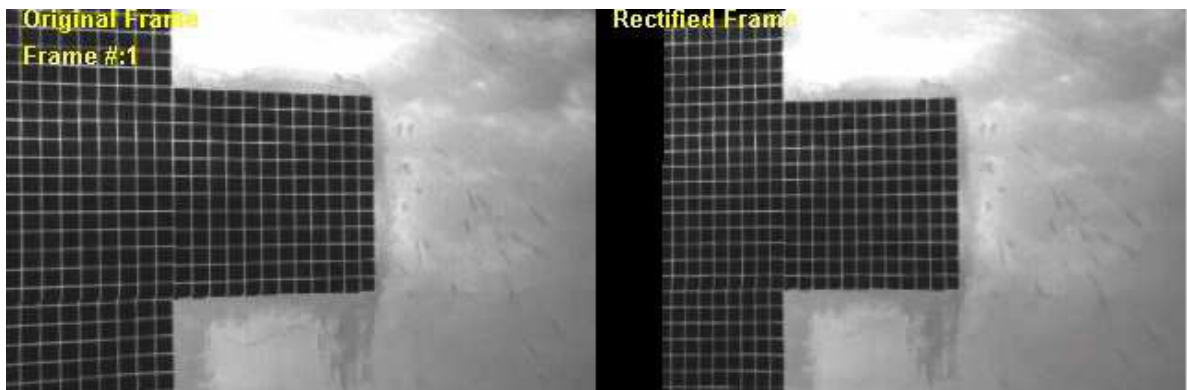
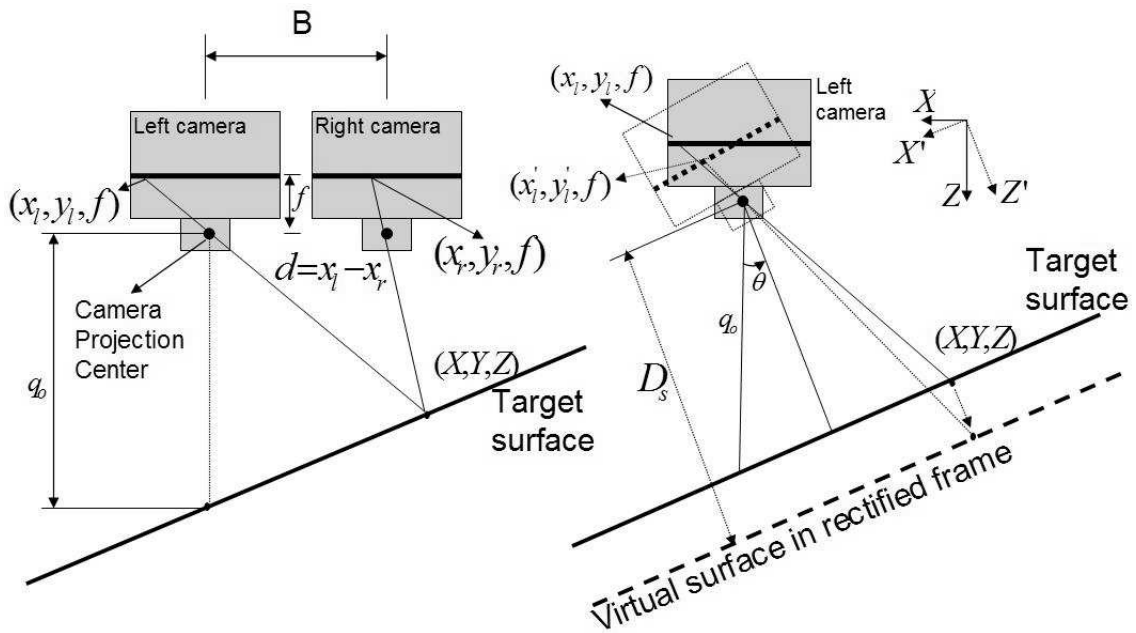


Fig. 3. Vehicle orientation and distance relative to the target, computed from disparity cues, can be used to generate a rectified video that simulates frontal views at fixed distance from the target. Each pixel in the rectified image is projected onto the rotated view based on position on the target surface (at desired position D_s), as depicted in the middle plot. The bottom image shows the pool wall surface viewed obliquely in the original image, and in the synthesized frontal view after rectification.

The solution, directly incorporating the orthogonality constraint $RRRR^T = I$, is given in terms of the left and right singular vectors of the 2×2 matrix $\hat{p}p_i^O(t+dt)\hat{p}p_i^O(t)^T$.

Integrating the displacements completes the measurements

of the ROV position, also enabling the construction of the mosaic from the video, say the left rectified sequence $I^l(t)$. Simultaneously, thruster control signals are generated based on the discrepancies from the desired heading and pitch angles,

and the distance to the hull at each time t .

IV. A DISCUSSION ON DRIFT ERROR ESTIMATION

As for most sensor-based navigation systems, drift error can play a significant factor in the performance of our vision-based navigation and mapping system. To arrive at a quantitative measure of drift rate, we need to account for a large number of variables, parameters, sources of noise, etc. These can vary drastically with environmental conditions, among other factors, and often cannot be readily estimated with accuracy particularly in a real-time operation. Therefore, estimating a realistic performance measure or drift rate is nontrivial. However, we will discuss some of the important factors, how they impact system performance, as well as how the drift characteristics of our system may be assessed if realistic estimates of the parameters from various factors are available.

We use the estimation uncertainty – namely, variance $\mathcal{V}_{\mathcal{E}}$ or covariance matrix $C_{\mathcal{E}}$ – to assess the propagation of error through various computational stages. In most cases, we have $\mathcal{V}_{\mathcal{E}_i} = (C_{\mathcal{E}}(i,i))^{1/2}$. As some computations involve nonlinear equations, we use the first-order approximation of these measures.

At the lowest level, the estimation of quantitative measures from visual cues depend on the image resolution, and in particular the precision in locating the image tokens (e.g., points) used in a particular computation. Assume that the uncertainty in localization an image point is represented in the form of a 2×2 covariance matrix C_p (for x and y components). Without loss of generality, we may assume $C_p = \mathcal{V}_p \mathbb{I}_{2 \times 2}$, where variance \mathcal{V}_p depends on a number of factors: texture characteristics and distance of the target scene, water turbidity conditions directly affecting image quality, accuracy of the feature detection algorithm, etc. This variance measure is the underlying critical factor in assessing the system performance. To establish how we may resolve image features with particular visual characteristics or detect their positions in underwater imagery, we can take advantage of ocean optics models and previous work on quantifying the image quality and contrast in terms of the water turbidity; e.g., [6], [25], [32], [33]. For example, assume that we determine \mathcal{V}_p based on a distinctness measure – say the corner measure of Harris interest point operator – attributed to each pixel [21]. We may further approximate the impact of the turbidity by a low-pass filtering operation, which directly affects the distinction measure. The filter parameter(s) – e.g., the variance of a Gaussian mask – can be parameterized in terms of the turbidity condition, target range, etc. As a result, we can arrive at the impact of the optical propagation path for a given turbidity condition on the image quality, contrast, distinctness measure, and thus the accuracy in localizing an image feature. We next examine how the uncertainty \mathcal{V}_p relates to a measure for drift error.

The first processing step involves establishing the orientation and distance of the ROV from the target, say a ship. Assuming the hull surface to be *locally* planar, disparities for a minimum of three points – giving us their positions on the surface – enables us to estimate the sought after information⁸.

Given that the disparity $dd = pp_i - pp_r$ is based on matching of pairs (pp_i, pp_r) in the left and right views, the variance \mathcal{V}_p of the 3-D point can be expressed in terms of \mathcal{V}_p : $\mathcal{V}_p = f_1(pp_i, pp_r, \mathcal{V}_p)$, where f_1 depends on the particular 3-D reconstruction method [22]. The computation involves covariance propagation according to $C_p = (J_p^p) C_p (J_p^p)^T$, where J_p^p is the Jacobian of the transformation from the image measurements to the 3-D scene point [20]. Consequently, we compute the uncertainty in the estimated vehicle orientation and distance to the target. For example, three points $PP_i = [X_i; Y_i; Z_i]$ ($i = 1 : 3$) give the equation of the local surface patch $Z = q_0 + q_1 X + q_2 Y$ from linear equations. Accuracy can be improved with the use of more points in a least-square formulation, preferably in combination with some outlier rejection method, e.g. RANSAC [12]. Parameters q_1 and q_2 establish the orientation and q_0 gives the distance. Furthermore, the estimation variance \mathcal{V}_{q_i} ($i = 1 : 3$) can be calculated from a model of the form $\mathcal{V}_{q_i} = f_2(X_i; Y_i; Z_i; \mathcal{V}_p)$. Again, f_2 depends on the estimation method, and is tied to $C_q = (J_q^q) C_q (J_q^q)^T$, where J_q^q is the Jacobian of the solution q_i 's in terms of the 3-D scene points P_i . We reiterate that these processing steps give the instant orientation and distance of the vehicle relative to the target, and do not require integration over time.

In the next step, the frontal views are constructed by image warping that can be expressed by the projective homography H_p , mapping points with homogeneous coordinates $\hat{p} = [x; y; z; 1]^T$ from original view onto points $\hat{p}^0 = [x^0; y^0; z^0; 1]^T$ in the frontal view: $\hat{p}^0 = H_p \hat{p}$. The 3×3 homography $H_p = RR + TT_z n n^T$ involves the rotation matrix RR (rotation by angle μ about axis n), translation TT_z (displacement $D_s = q_0 \cos \mu$ along the viewing Z direction to maintain the distance D_s to the wall), and surface normal $n n^T = (q_1; q_2; 1) = q_0$; see discussion in the last section.

The variances \mathcal{V}_{q_i} 's directly establish the variances \mathcal{V}_{h_i} 's of the homography parameters h_i 's: $\mathcal{V}_{h_i} = f_3(q_i; \mathcal{V}_{q_i})$. Again, we utilize the homography model and the Jacobian with respect to q_i 's: $C_h = (J_h^q) C_q (J_h^q)^T$ (note that H_p can be expressed in terms of q_i ($i = 1 : 3$)). Consequently, we compute the uncertainty in the position of each pixel in the warped image (synthesized frontal view) from $\mathcal{V}_{p^0} = f_4(h_i; pp_i; \mathcal{V}_{h_i}; \mathcal{V}_{q_i})$ [5], where f_4 is established by the projective homography and the covariance propagation: $C_{p^0} = (J_{p^0}^p) C_p (J_{p^0}^p)^T + (J_{p^0}^h) C_h (J_{p^0}^h)^T$. The interpolation process in the construction of the frontal view contributes additional uncertainty. However, this is often negligible compared to the much larger uncertainty given above.

The next step involves accounting for the estimation uncertainty of the planar motions from one ROV position to the next, either based on the 2-D correlation scheme, or the isotropic/similarity transformation model. In the first case, the estimation uncertainty is trivially determined from the weighted sum of the pixel uncertainties used in the computation. In the two other cases, we can utilize the theoretical uncertainty measures derived for the absolute orientation problem [9], simplified to two dimensions for the isotropic/similarity transformation between each pair of consecutive images. Generally, the result can be expressed in the form $\mathcal{V}_{q_i} =$

⁸Alternatively, we can correlate certain local regions in the two views.

$f_5(pp^0(t+dt); pp^0(t); \frac{3}{4}p^0)$, where f_5 is tied to the closed-form solution for the similarity transformation parameters, and $\frac{3}{4}p^0$'s give the variances of the transformation parameters (translations t_x and t_y , rotation t_z , and possibly scales).

Having derived the variances of the motion components, we are ready to assess the damage from the integration over time (drift error) of the side-way motions that are estimated from pairs of consecutive synthesized frontal views. Skipping the scale factor s , this can be readily done either for the 3-D ROV position PP_R based on the equation [9],

$$PP_R(t+dt) = RR(t)PP_R(t) + TT(t);$$

or the 2-D mosaic points according to the 2-D similarity transformation

$$pp^0(t+dt) = RR(t)pp^0(t) + TT_{xy};$$

Given the similarity between the two equations, we give the simpler expression for the covariance of the mosaic points:

$$C_{pp^0(t+dt)} = J_t C_{W(t)} J_t^T;$$

where the covariance $C_{W(t)}$ of the motion and position vector $w(t) = [t_z(t); T_x(t); T_y(t); x^0(t); y^0(t)]$ is given by

$$C_{W(t)} = \begin{bmatrix} C_{t_z; T_x; T_y} & \\ & C_{pp^0(t)} \end{bmatrix};$$

and the Jacobian J of the transformation can be written

$$J_t = \frac{\partial pp^0(t+dt)}{\partial w(t)} = \begin{bmatrix} x^0(t) & y^0(t) & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{bmatrix} \begin{bmatrix} RR(t) \\ & RR(t) \\ & & RR(t) \\ & & & RR(t) \\ & & & & RR(t) \end{bmatrix};$$

The process is initialized with $C_{pp^0(0)} = 0_{2 \times 2}$.

Ending the dry mathematical derivation of the drift error on a more positive note, it is appropriate to comment on ways that it may be reduced or eliminated. Where feasible, this can be achieved by the integration with an absolute positioning system. As an example, an existing product utilizes the LBL acoustic technology with a map of the ship to determine regions of the hull that have been inspected by the ROV/diver [53]. Integration with the stereovision technology enables painting the map with a mosaic, incorporating the 3-D target detection capabilities, and potentially improving the localization accuracy by fusing high-resolution visual servoing with drift-free acoustic-based estimates.

V. EXPERIMENTS

The ship inspection system integrates three automated real-time capabilities solely from stereo imagery: Estimation of positioning and navigational information, construction of 2-D photo-mosaics and computation of 3-D disparity maps that encode the 3-D target shape. Fig. 4 (top) depicts the system display screen summarizing these various capabilities: The left image of the stereo pair shown in top left section, and the corresponding reconstructed disparity as a color map right below. The live mosaic is given to the right, with various motion components above. In addition, instantaneous

positioning and motion information –distance and orientation to the target and the xy motion are displayed directly on the live image.

Experiments with the ROV has been carried out both in the pool and at a dock in open waters; see fig. 4 (bottom). Two sample experiments are presented to demonstrate various system functionalities, and its performance. In particular, the results from the pool tests are compared with those based on the use of monocular cues in order to demonstrate some of the main advantages in the use of a stereovision system.

In these tests, the baseline was fixed at 5 inches, which is suitable the 3-5 ft operational range. The cameras have been calibrated, both for internal and external parameters[23]. The system runs on a 2.4MHz Dual Xeon processor, and carries out both the stereo and motion computations at 15 Hz.

A. Pool Test

Pool tests have been carried out to highlight some important capabilities of the system. Specifically, they allow us to determine if acceptable performance can be achieved when the target surface is relatively textureless, and one may expect the most significant adverse impact from shading and shadow artifacts due to surface waves (see also fig. 2). At the same time, we can verify if these cast shadows (in addition to the actual surface texture) can truly be exploited as projected markings for region matching in stereo images, rather than posing a serious challenge. In the experiment described here, the ROV is positioned at a distance of about 5 ft from the pool wall; fig. 4 (left). It is manually pulled parallel to the wall, while performing a sinusoidal-type motion, both in the Z and heading directions. The pitch angle is varied similarly, but with lower variations (since the vehicle is floating near the surface). The vertical motion is limited, due to shallow (4 ft) depth of the pool. Fig. 5 gives a sample image of the wall, the estimated pose (in degrees) relative to the pool, and the position along the path. The estimated side-way distance traveled is roughly 7 meters, and aside from the minimal roll motion, all other significant motion components are estimated with good accuracy. This is also verified from the constructed mosaic which agrees with the manual measurement of the distance covered, and the locations of light fixtures and a fan. During various system developing and testing phases, numerous (more than 2 dozens) runs were done where the ROV was instructed to travel the length of the 25-yard pool while maintaining a fixed distance and heading. In these cases, the ROV started and ended at different, but nearby, spots at the two ends with varying stand-off distances of 50-120 cm. Though the exact covered distance was not measured manually, observed estimates typically varied in the range of 23.5 to 24.25 yards, which were within the experimental accuracy (0.5 yard average error would correspond to 2% of the traveled distance).

Fig. 6 depict various mosaics generated by processing only the left sequence of the stereo video data, based on various image transformation models. All of these mosaics have been rendered in the coordinate frame of the first image, as with one constructed with stereo cues. The immediate

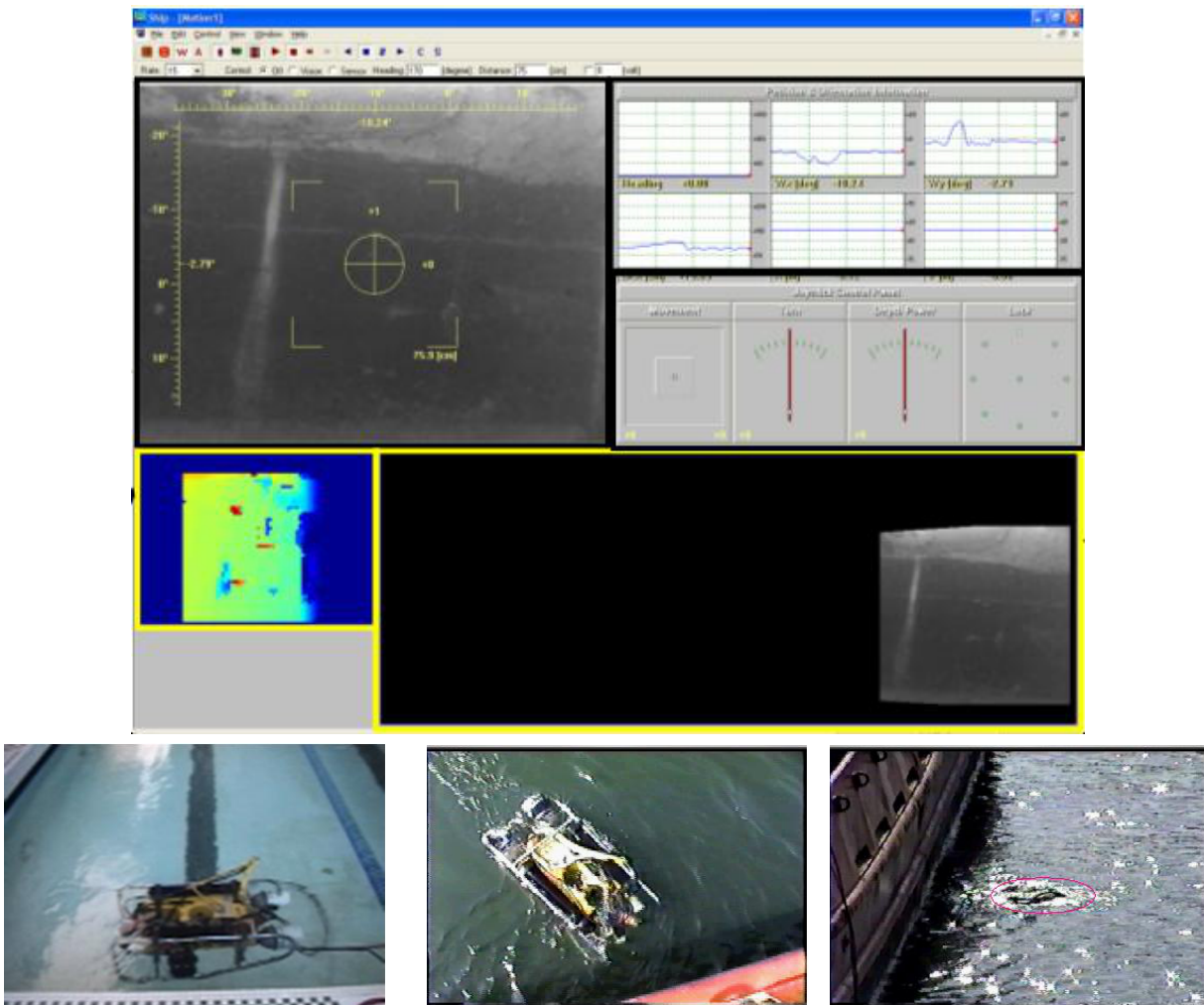


Fig. 4. Top: Inspection system's computer display with the live (left) view, superimposed with various instantaneous positioning and motion information, disparity as a color map, 2-D photo-mosaic, and 6 components of position information as time plots (see text for details). Bottom: Live shots of ROV during demos in the pool (left) and the dock test (middle and right).

observation is that not only the drift error grows relatively fast, the performance deteriorates with the increasing d.o.f. in the transformation model. In particular, while the projective model is the most suitable transformation as the ROV motion involves all 6 d.o.f., it gives the worst performance (highest distortion) among various methods. This is directly tied to the fact that, with the low-texture characteristics of the scene, one cannot reliably extract and match as many features that are typically required to estimate the transformation parameters with good accuracy (see below). Furthermore, the camera tilt with respect to the pool wall in the first frame results in later frames to be projected "beyond infinity" in the mosaic frame. Also, one notes from images 53 and 54 (bottom-right image pair) that the ROV appears to be slightly tilted downwards. We can differentiate whether this is an instantaneous orientation of the ROV or that the cameras are mounted with some small-angle tilt. Noting that the mosaic based on the similarity transform depicts a curved strip with center of curvature below, we can establish the small-angle downward tilt of the cameras. These various so-called "misalignments" between the camera coordinate system and the wall surface are rather difficult to determine from monocular cues. In contrast, they have been

readily estimated and thus accounted for in construction the mosaic by processing of the stereo data.

The bottom row shows two sample consecutive images, with the detected features in each pair. Inlier matches have been shown by green crosses, while the red crosses depict the outlier matches. These are worse and typical best case scenarios in processing the entire data set. In particular, four correct matches is the minimum necessary to compute the projective homography. Furthermore, the fact that these features cover a rather small camera f.o.v. is a major factor in the non-robust estimation of the transformation parameters in the models with a larger number of degrees of freedom. We reiterate once again another important advantage of binocular cues that is particularly relevant when dealing with such pool-wall type low-textured targets. In processing the monocular data, many weak surface markings are not matched because of the local variations in visual appearance in consecutive frames, e.g., due to the more dominant time-varying shading and cast shadows from the surface waves. To the extent that these affect the left and right views similarly, the features can be matched in stereo data. While the temporal matching difficulty exists, only a small number (2 in theory) of correct matches are necessary

as the estimation is restricted to image plane motions (up to 3, out of 6, degrees of freedom) in the rectified video.

B. Dock Test

Several dock tests have also been performed at the Sea Technology Center of Florida Atlantic University, Dania Beach, FL; see fig. 4 (bottom). In these tests, the ROV was instructed to travel parallel to a ship hull at the dock, while maintaining a fixed distance (about 3 feet) and heading (perpendicular direction to the hull). The system operated automatically (under computer control) based solely on the position information from stereo imagery, while exposed to random wave disturbances from boats travelling through the entrance of the nearby waterway. Fig. 5 depicts the results from a typical test. It comprising the first frame of the video at one end of the hull, and estimates of pose and position of the ROV during the test. In the next plot, the reconstruction consistency is demonstrated by comparing the corresponding mosaic with those from two other runs. In particular, geometric accuracy can be assessed through comparison between any one mosaic and each of selected frames of the video, given in the last two rows.

Finally, fig. 8 shows selected left views from a trial to test the target detection capability. The disparity (inverse target distance) below each image is depicted as a color map (blue is far, and red is close range). Roughly, only the region marked by the red rectangle (on the first image) is within the f.o.v. of both cameras, and thus can be processed for 3-D reconstruction. To provide a true assessment of the tested processing capabilities, we have given here results from a real-time operation, without any post processing to enhance performance. In particular, no filtering step had been incorporated to remove isolated noisy estimates or to improve estimates at the edges, to maintain 15Hz processing at the time of this demonstration. The disparity computation, carried out over 12×8 windows to reduce sensitivity to floating particles, clearly results in extended blurring at the object boundaries, as noted. However, one can clearly discern the spherical object, the step, and ship surface boundaries.

VI. SUMMARY

Inspections of ship hulls, bridges, port docks, dams and similar structures are suitable applications of ROVs. Effectiveness of the operation is directly tied to capabilities that can be carried out automatically, including navigation and mapping, with or without human interaction and supervision. This paper addresses the visual inspection, navigation and mapping solely based on optical imaging. While the challenges of this task by utilizing a monocular system may be hard to overcome, we have demonstrated the significant benefits of stereo vision for the realization of a robust system. In particular, we have highlighted the inherent complexities faced by a monocular system, and how they are resolved with binocular vision.

A real-time system has been developed which, running on a 2.4 MHz dual Xeon processor, carries out stereo and motion computations at 15 Hz. The system has been tested in a number of experiments in a pool and open sea. The pool

experiments have demonstrated the successful performance of the system in the absence of surface texture, which often poses serious problems for feature matching in stereo and motion. Moreover, being able to control some aspects of the experiment, including some rough knowledge of ground truth, has been critical during developmental stages and in improving the performance. The dock tests, providing assessment of the system performance under more realistic uncontrolled conditions, clearly demonstrate the potential in the deployment of this technology.

Acknowledgement We are in debt to Nuno Gracias who processed the pool data with his automated system to construct mosaics from monocular cues [19]. We are thankful to the reviewers whose comments allowed us to significantly improve the paper from the first draft. This article is based upon work supported by the Office of Naval Research (ONR) under Grant No. N000140310074, recommended by Dr. Thomas Swain. We are grateful to Mr. Chris O'Donnell and David Gill (NEODTD, Indian Head, MD) for their support in getting the project off the ground, Ms. Penny Lanham (NEODTD) for valuable advice throughout the project and particularly discussions and recommendation (with Mr. Andy Pedersen, NEODTD) at the first live demo, and finally Mr. Craig Lawrence and his crew of Navy divers who made the long trip down to Panama City to help us in video data collection. We sincerely thank Mr. Alan Rose and his staff at the UoM Wellness Center for the unlimited use of the indoor pool facility, and making every effort to accommodate us, both through the system development and testing phases, and for our first project live demo. We are in great debt to our dear FAU friends and colleagues, particularly Professor Cuschieri of the Ocean Engineering Department (now with Lockheed Perry), for allowing us the use of the Sea Technology Center's dock facility. Any opinions, findings, conclusions or recommendations expressed in this manuscript are those of the authors and do not necessarily reflect the views of the ONR and (or) NEODTD.

REFERENCES

- [1] Adiv, G., "Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field," *IEEE Trans. PAMI*, Vol 11(5), May, 1986.
- [2] Balasuriya, A., T. Ura, "Underwater cable following by Twin-Burger 2," *Proc. IEEE Robotics Automation*, 2001.
- [3] Barufaldi, C., P. Firoozfam, and S. Negahdaripour, "An integrated vision-based positioning system for video stabilization and accurate local navigation and terrain mapping," *Proc. Oceans'03*, San Diego, September 2003.
- [4] J. Battle, E. Mouaddib, and J. Salvi, "Recent progress in coded structured light as a technique to solve the correspondence problem: a survey," *Pattern Recognition*, 31(7), 1998.
- [5] Criminisi, A., I. Reid, and A. Zisserman, "A plane measuring device," *Image Vision and Computing*, 17, 1999.
- [6] Duntley, S.Q., "Light in the sea," *Journal of the Optical Society of America*, 53, 1963.
- [7] Eustice, R., H. Singh, J. Howland, "Image registration underwater for fluid flow measurements and photomosaicking," *Proc. IEEE/MTS Oceans*, Providence, RI, 2000.
- [8] Fan, Y., and A. Balasuriya, "Optical flow based speed estimation in AUV target tracking," *Proc. Oceans*, Honolulu, HI, 2001.
- [9] Firoozfam, P., and S. Negahdaripour, "Theoretical accuracy analysis of N-ocular vision systems for scene reconstruction, motion estimation, and positioning," *Proc. Int. Conf. 3-D Processing, Visualization and Transmission*, Thessaloniki, Greece, September, 2004.

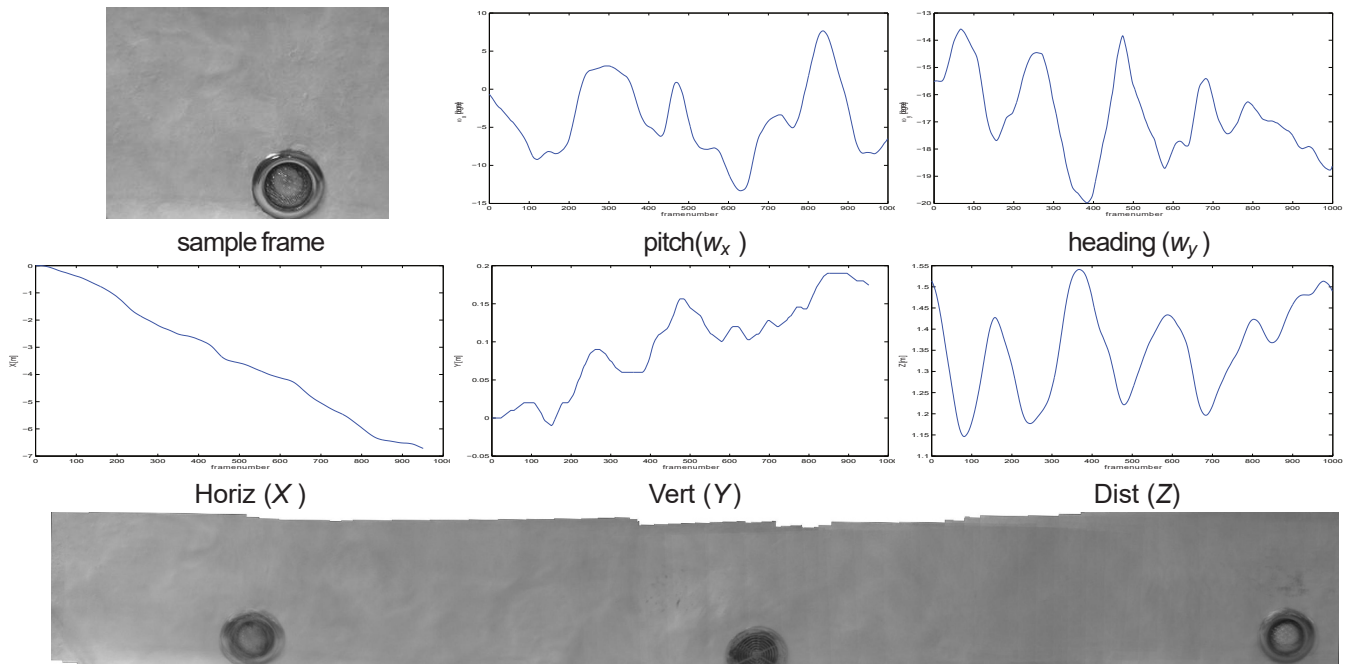


Fig. 5. Pool test results; ROV positioning and mosaic construction based of stereo imagery. Estimated ROV pose (pitch and heading in degrees) is relative to the pool wall, X Y distances are relative to the start position, and Z is the distance from the wall.

- [10] Fleischer, S.D., H.W. Wang, S.M. Rock, M.J. Lee, "Video mosaicking along arbitrary vehicle paths," Proc. Symp. AUV Technology, Monterey, CA, June, 1996.
- [11] Fleischer, S.D., S.M. Rock, "Global position determination and vehicle path estimation from a vision sensor for real-time mosaicking and navigation," Proc. Oceans, Halifax, Nova Scotia, October, 1997.
- [12] Fischler, M.A. and R.C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," Comm ACM, Vol 2(6), 1981.
- [13] Fusiello, A., E. Trucco, and A. Verri, "A compact algorithm for rectification of stereo pairs," Machine Vision and Applications, Vol 12, 2000.
- [14] Garcia, R., X. Cufí, J. Batlle, "A system to evaluate the accuracy of a visual mosaicking methodology," Proc. Oceans'01, Honolulu, HI, November, 2001.
- [15] Garcia, R., X. Cufí and V. Ila, "Recovering camera motion in a sequence of underwater images through mosaicking," Lecture Notes in Computer Science, No. 2652, Springer-Verlag, 2003.
- [16] Garcia, R., T. Nicosevici, P. Ridao, and D. Ribas, "Towards a real-time vision-based navigation system for a small-class UUV," IEEE IROS, Las Vegas, 2003.
- [17] Gracias, N., J. Santos-Victor, "Underwater video mosaics as visual navigation maps," Computer Vision Image Understanding, Vol 79, July, 2000.
- [18] Gracias, N., J. Santos-Victor, "Underwater mosaicking and trajectory reconstruction using global alignment," Proc. Oceans'01, Honolulu, HI, November, 2001.
- [19] Gracias, N., "Mosaic-based visual navigation for autonomous underwater vehicles," Ph.D. Thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisbon, Portugal, June, 2003.
- [20] Haralick, R.M. "Propagating covariance in computer vision," ICPR-A, 1994.
- [21] Harris, C. J., and M. Stephens. "A combined corner and edge detector," Proc. 4th Alvey Vision Conference, Manchester, 1988.
- [22] Hartley, R.I., and A. Zisserman, "Multiple view geometry in computer vision," Cambridge Press, 2000.
- [23] Heikkilä, J. and O. Silvén, "A four-step camera calibration procedure with implicit image correction," Proc. CVPR, Puerto Rico, 1997.
- [24] Horn, B.K.P., M. Hilden, and S. Negahdaripour, "Closed form solutions of absolute orientation using orthonormal matrices," JOSA-A, Vol 5(7), 1987.
- [25] Jaffe, J.S., "Computer modeling and the design of optimal underwater imaging systems," IEEE Journal of Oceanic Engineering, Vol 15, No. 2, April, 1990.
- [26] Marchand, E., F. Chaumette, F. Spindler, M. Perrier. "Controlling the manipulator of an underwater ROV using a coarse calibrated pan tilt camera," IEEE Int. Conf. Robotics Automation, Volume 3, Seoul, S. Korea, May, 2001.
- [27] Marks, R., H.H. Wang, M. Lee, and S. Rock, "Automatic visual station keeping of an underwater robot," Proc. Proc. Oceans'94, Brest, France, September, 1994.
- [28] Leabourne, K.N., S.M. Rock, S.D. Fleischer, and R.L. Burton. "Station keeping of an ROV using vision technology," Proceedings OCEANS, Halifax, Nova Scotia, October 1997.
- [29] Lots, J-F, D. M. Lane, E. Trucco, and F. Chaumette, "A 2-D visual servoing for underwater vehicle station keeping," Proceedings IEEE Conference Robotics and Automation, Seoul, Korea, 2001.
- [30] Marks, R., H.H. Wang, M. Lee, and S. Rock, "Automatic visual station keeping of an underwater robot," Proc. Proc. Oceans'94, Brest, France, September, 1994.
- [31] Majumder, S., S. Scheduling, and H.F. Durrant-Whyte, "Multi sensor data fusion for underwater navigation," Robotics and Autonomous Systems, Vol 35(1), 2001.
- [32] McGlamery, B.L., "Computer analysis and simulation of underwater camera system performance," SIO Ref. 75-2, Scripps Inst. of Oceanography, UCSD, January, 1975.
- [33] Mobley, Handbook of Optics, Bass, M. (Editor), Second Edition, Optical Society of America and McGraw-Hill, 1992.
- [34] Negahdaripour, S., and J. Fox, "Underwater optical station keeping: Improved methods," Journal of Robotic Systems, March, 1991.
- [35] Negahdaripour, S., A. Shokrollahi, C.H. Yu "Optical Sensing for underwater robotic vehicles," Journal of Robotics and Autonomous Systems, Vol 7, 1991.
- [36] Negahdaripour, S., X. Xun, A. Khamene, "Applications of direct 3D motion estimation for underwater machine vision systems," Proc. Oceans'98, Nice, France, September, 1998.
- [37] Negahdaripour, S., X. Xu, L. Jin, "Direct estimation of motion from seafloor images for automatic station-keeping of submersible platforms," IEEE Journal Oceanic Eng., Vol 24(3), July, 1999.
- [38] Negahdaripour, S., X. Xu, "Mosaic-based positioning and improved motion estimation methods for autonomous navigation," IEEE Journal Oceanic Engineering, Vol 27 (1), January, 2002.
- [39] Ortiz, A., G. Oliver, and J. Frau, "A vision system for underwater real-time control tasks," Proc. Oceans'97, Halifax, CA, Oct. 1997.
- [40] C. Plakas, E. Trucco, A. Fusiello, "Uncalibrated vision for 3-D underwater applications," Proc. Oceans'98, Nice, France, September, 1998.
- [41] Roman, C., and H. Singh, "Estimation of error in large area underwater

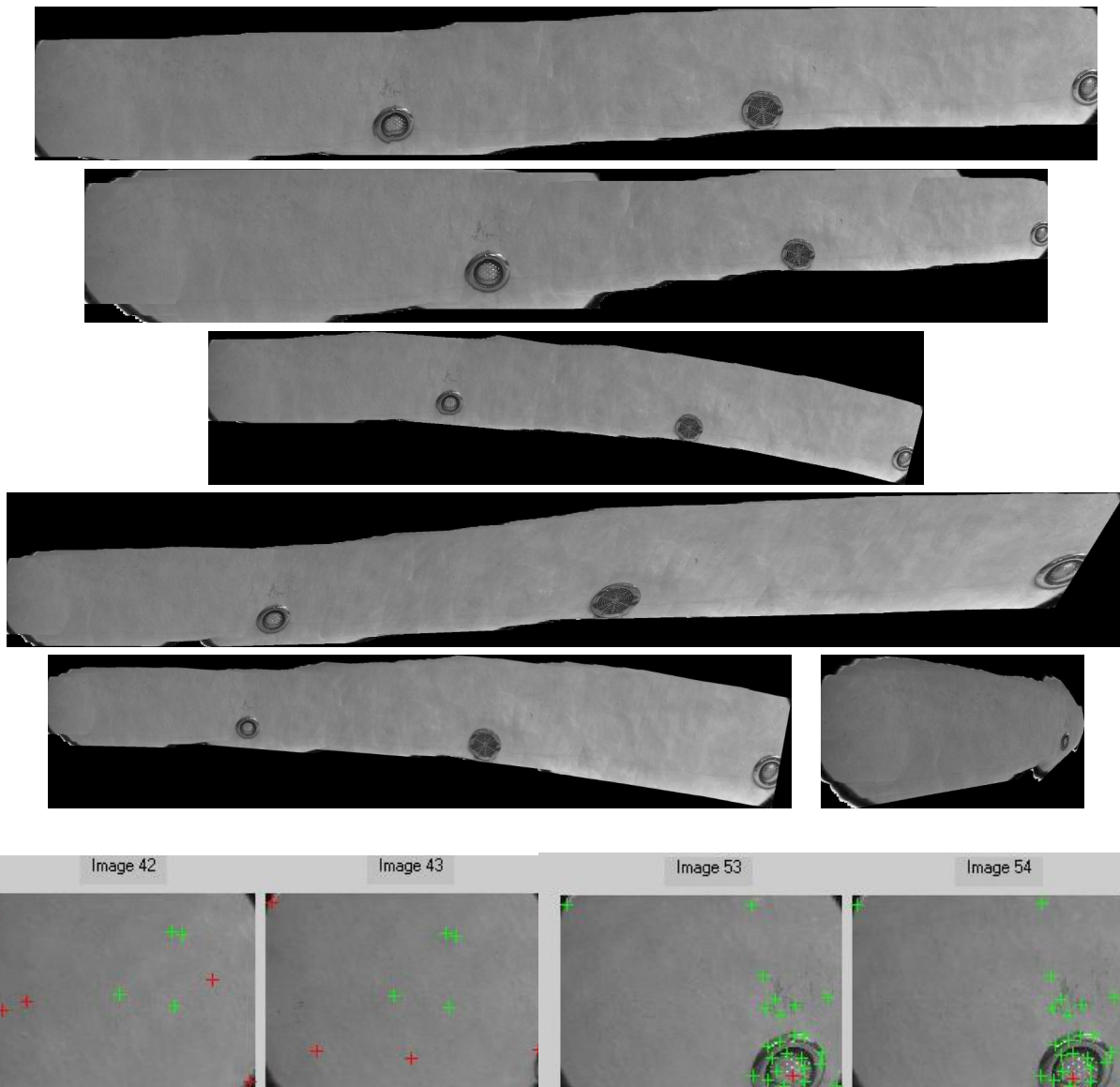


Fig. 6. Pool data mosaics from monocular cues (left sequence of the stereo video) based on various image transformation models. From top: Planar translation; translation and zoom; planar translation and rotation; affine model; similarity transformation; planar (projective) homography. Bottom row shows two sample sequences with detected inlier (green) and outlier (red) matches. See text for details.

- photomosaics using vehicle navigation data," *Proc. Oceans*, Honolulu, HI, November, 2001.
- [42] Rosen, D., "ROV role in homeland security," *Proc. Underwater Intervention*, 2003.
- [43] Rzhanov, Y., L.M. Linnett, R. Forbes, "Underwater video mosaicking from seabed mapping," *Proc. Int. Conf. Image Proc.*, Vancouver, CA, 2000.
- [44] G. Sansoni, M. Carocci, and R. Rodella, "Calibration and performance evaluation of a 3D imaging sensor based on the projection of structured light," *IEEE Tran. Instrumentation and Measurement*, 49(3), June 2000.
- [45] Silpa-Anan, C., T. Brinsmead, S. Abdallah, and A. Zelinsky "Preliminary experiments in visual servo control for autonomous underwater vehicle," *Proc. IEEE Int. Conf. Intelligent Robots and Systems*, Maui, Hawaii, October/November 2001.
- [46] Trucco, E., A. Doull, F. Odone, A. Fusiello, and D. M. Lane, "Dynamic video mosaics and augmented reality for subseaspection and monitoring," *Oceanology Int.*, Brighton, England, 2000.
- [47] Vincent A.G., N. Pessel, M. Borgetto, J. Jouffroy, J. Operbecke and V. Rigaud, "Real-time geo-referenced video mosaicking with the MATISSE system," *Proc. Oceans'03*, San Diego, CA, September, 2003.
- [48] Wettergreen, D., C. Gaskett, and A. Zelinsky "Development of a visually-guided autonomous underwater vehicle," *Proc. Oceans'98*, Nice, France, September, 1998.
- [49] Xun, S. Negahdaripour, "Vision-based motion sensing for underwater navigation and mosaicing of ocean floor images," *Proc. Oceans*, Halifax, Canada, October, 1997.
- [50] *Proc. US Navy and Coast Guard Workshop on Hull Search*, USF Ocean Tech Center, St. Petersburg, FL, April, 2003.
- [51] Z.Y. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," *Proc. IEEE Int. Conf. Computer Vision*, 1999.
- [52] http://eia.udg.es/~jpages/coded_light/
- [53] <http://www.desertstar.com/newsite/positioning/ship/hull/ship.html>

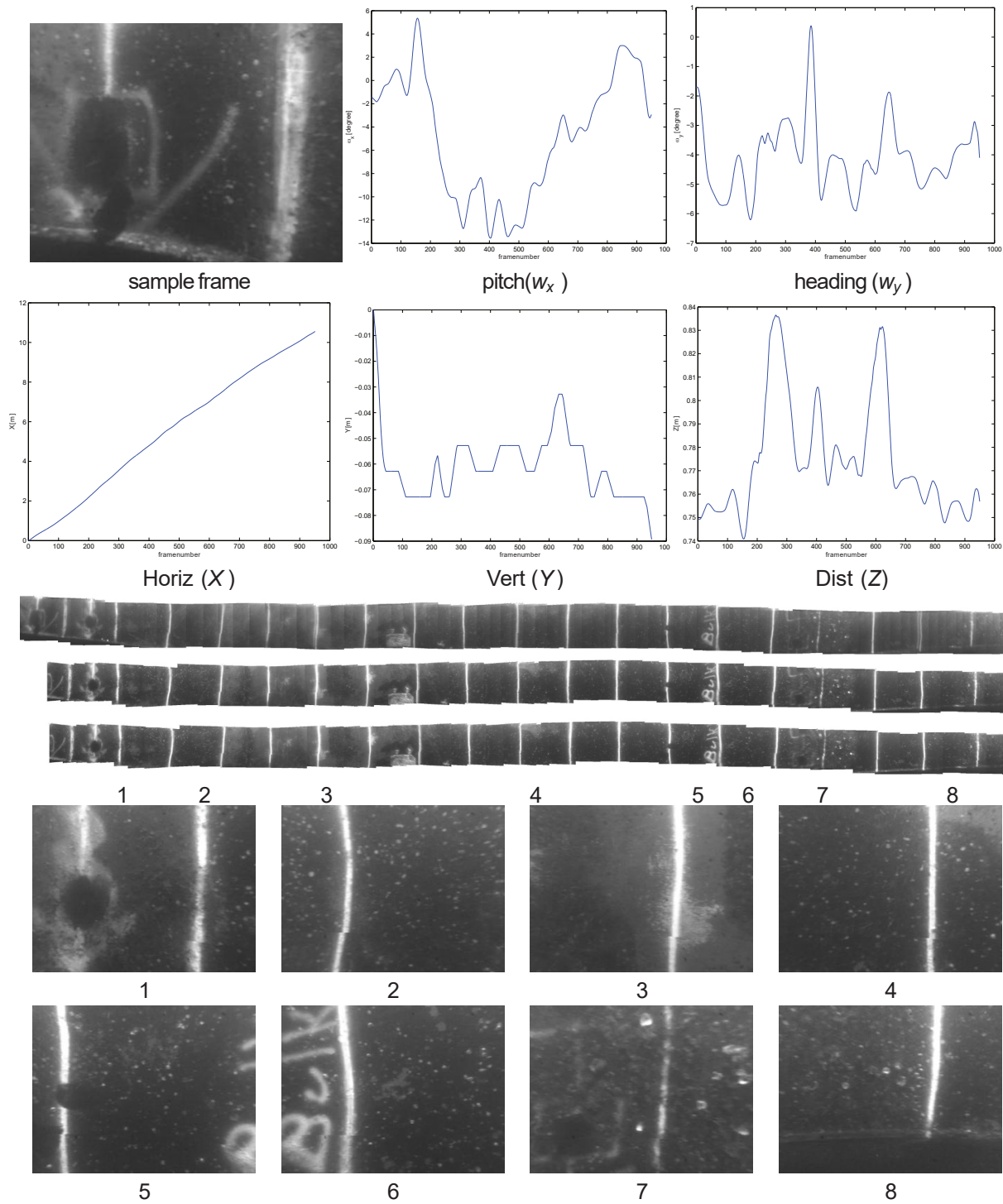


Fig. 7. Dock Test; Automatic ROV positioning in constructing the mosaic of a ship hull based of stereo imagery. First frame, ROV pose (pitch and heading in degrees) relative to the ship, and the X Y distances (in meters) relative to the start position, and distance Z (in meters) from the hull are given. Next plot is the corresponding mosaic and two more from different runs given for comparison to demonstrate consistency. Geometric accuracy can be verified by noting selected frames of the left video, given in the last two rows, used in construction the mosaic.



Shahriar Negahdaripour received S.B., S.M., and Ph.D. degrees in 1979, 1980, and 1987 from Massachusetts Institute of Technology, Cambridge. He joined the Electrical Engineering Department, University of Hawaii, Honolulu, in January 1987 as an Assistant Professor. In August 1991, he joined the University of Miami, Coral Gables, FL, where he is currently a Professor

of electrical and computer engineering. Since the start of a project on automatic vision- based ROV station keeping in 1988, supported by the University of Hawaii Sea Grant College Program, he has been involved in a number of projects on the development of various vision technologies for underwater applications, primarily supported by the National Science Foundation, Office of Naval Research, NAVSEA (Indian Head, MD) and Naval Undersea Warfare Center (Newport, RI), and NOAA (Ocean Exploration Program). In addition

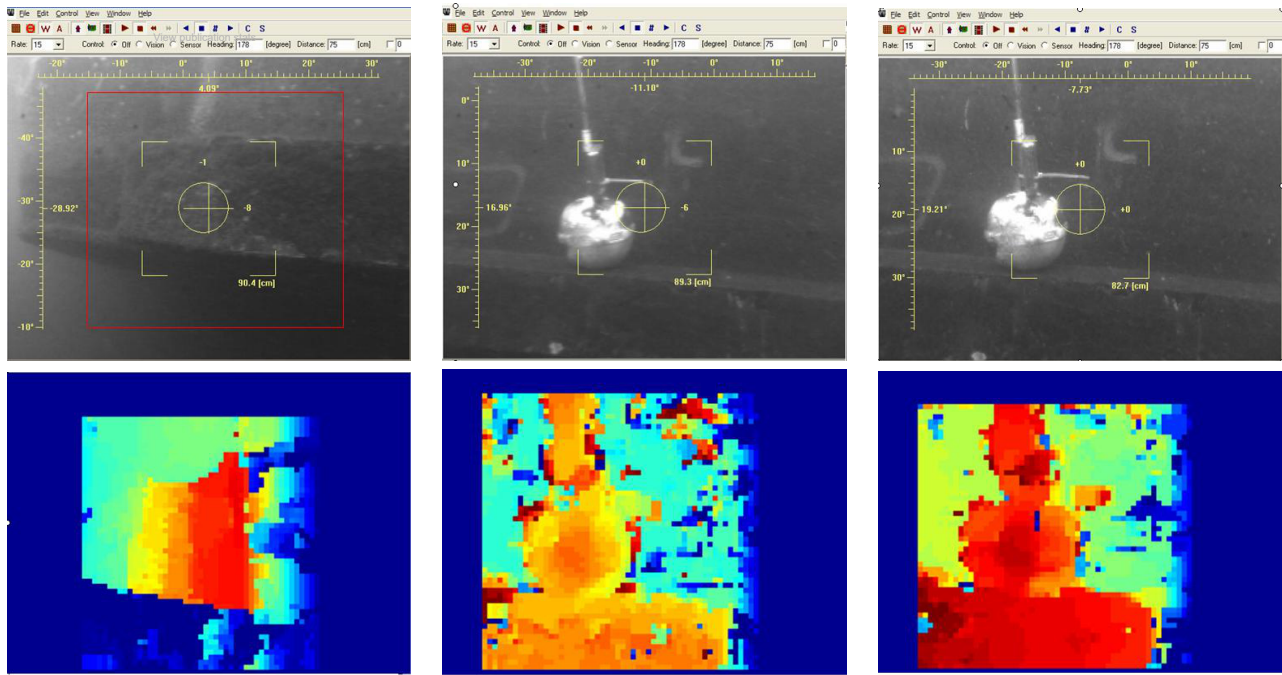


Fig. 8. Dock test of real-time inspection system; Selected left view of a stereo pair, and computed stereo disparity as a color map (blue far and red close range).

to numerous journal and conference papers in this subject area, he has presented a half-day tutorial at the IEEE OCEANS'98 Conference, Nice, France, on "Computer Vision for Underwater Applications." Dr. Negahdaripour has organized and moderated panel sessions at IEEE OCEANS'01 and OCEANS'02 conferences on "Robust Video Mosaicking: Techniques, Standards, and Data Sets," and "Photo-Mosaicking of Underwater Imagery: Challenges and Technical Issues."



Pezhman Firoozfam received B.Sc. degree in Electrical/Electronics Engineering from KNT University of Technology, Tehran, Iran, in 1994, M.Sc. degree in Bio-electrical Engineering from Sharif University of Technology, Tehran, Iran, in 1997, and Ph.D. degree in Electrical and Computer Engineering from the University of Miami, Coral Gables, FL in 2004. His research interests include 3-D motion estimation and scene reconstruction from N-Ocular and panoramic imagery for sea-floor and underwater structures, as well as close-range imaging and computer vision.

He has also been involved in different areas of geomatics engineering and photogrammetry for several years.