

# Fourier-based Registration for Robust Forward-looking Sonar Mosaicing in Low-visibility Underwater Environments



## Natàlia Hurtós

Computer Vision and Robotics Group, University of Girona, Girona 17071, Spain  
e-mail: [nhurtos@eia.udg.edu](mailto:nhurtos@eia.udg.edu)

## David Ribas

Computer Vision and Robotics Group, University of Girona, Girona 17071, Spain  
e-mail: [dribas@eia.udg.edu](mailto:dribas@eia.udg.edu)

## Xavier Cufi

Computer Vision and Robotics Group, University of Girona, Girona 17071, Spain  
e-mail: [xcuf@eia.udg.edu](mailto:xcuf@eia.udg.edu)

## Yvan Petillot

Ocean Systems Laboratory, Heriot-Watt University, EH14 4AS Edinburgh, United Kingdom  
e-mail: [y.r.petillot@hw.ac.uk](mailto:y.r.petillot@hw.ac.uk)

## Joaquim Salvi

Computer Vision and Robotics Group, University of Girona, Girona 17071, Spain  
e-mail: [jsalvi@eia.udg.edu](mailto:jsalvi@eia.udg.edu)

Received 31 December 2013; accepted 13 February 2014

Vehicle operations in underwater environments are often compromised by poor visibility conditions. For instance, the perception range of optical devices is heavily constrained in turbid waters, thus complicating navigation and mapping tasks in environments such as harbors, bays, or rivers. A new generation of high-definition forward-looking sonars providing acoustic imagery at high frame rates has recently emerged as a promising alternative for working under these challenging conditions. However, the characteristics of the sonar data introduce difficulties in image registration, a key step in mosaicing and motion estimation applications. In this work, we propose the use of a Fourier-based registration technique capable of handling the low resolution, noise, and artifacts associated with sonar image formation. When compared to a state-of-the-art region-based technique, our approach shows superior performance in the alignment of both consecutive and nonconsecutive views as well as higher robustness in featureless environments. The method is used to compute pose constraints between sonar frames that, integrated inside a global alignment framework, enable the rendering of consistent acoustic mosaics with high detail and increased resolution. An extensive experimental section is reported showing results in relevant field applications, such as ship hull inspection and harbor mapping. © 2014 Wiley Periodicals, Inc.

## 1. INTRODUCTION

Over the past few years, underwater vehicles have greatly improved as a tool for undersea exploration. In particular, autonomous navigation, localization, and mapping through optical imaging have become topics of interest for researchers in both underwater robotics and marine science communities. Underwater imagery has been used to construct image photomosaics with applications in harbor security (Negahdaripour & Firoozfam, 2006), environmental monitoring (Elibol et al., 2011), and damage assessment (Lirman et al., 2010), being a key tool to locate areas or ob-

jects of interest, detect changes, or plan subsequent missions in an area. Likewise, underwater navigation has benefited from visual processing methods such as visual odometry and visual simultaneous localization and mapping (SLAM) (Eustice, Pizarro, & Singh, 2008; Gracias, Van Der Zwaan, Bernardino, & Santos-Victor, 2003) to provide drift-free navigation using onboard cameras.

However, a significant number of surveying and mapping tasks in underwater scenarios are carried out in turbid waters and murky environments where vehicles equipped only with optical systems (i.e., cameras or lasers) are constrained by their limited visibility range. Knowing the

limitations of optical devices, underwater operations have long relied on sonar technology for obstacle avoidance, navigation, localization, and mapping (Fairfield, Jonak, Kantor, & Wettergreen, 2007; Kinsey, Eustice, & Whitcomb, 2006; Leonard, Bennett, Smith, & Feder, 1998; Roman & Singh, 2005; Tena, Reed, Petillot, Bell, & Lane, 2003) by employing different types of sonar (e.g., profiling sonar, multibeam echosounders, scanning imaging sonar, side-scan sonar). Recently, a new generation of sonars, namely the two-dimensional forward-looking sonars (2D FLS), have emerged as a strong alternative for those environments with reduced visibility given their capabilities of delivering high-definition acoustic images at a near-video frame rate (BlueView Technologies Inc., 2013; Soundmetrics Corp., 2013; Tritech Gemini, 2013).

Several researchers have drawn attention to the use of these high-frequency sonars either as a substitute or as a complementary device for optical cameras (Negahdaripour, Sekkati, & Pirsivash, 2009). FLS imagery has been employed in benthic habitat mapping (Negahdaripour et al., 2011), monitoring of fish populations (Baumgartner & Wales, 2006), detection of targets on the seafloor (Galceran, Djapic, Carreras, & Williams, 2012), and inspection of ship hulls (Hover et al., 2012). The integration of FLS in a visual SLAM framework to constrain the navigation drift of autonomous underwater vehicles (AUVs) has also been a topic of interest (Johannsson, Kaess, Englot, Hover, & Leonard, 2010; Walter, 2008).

The processing of FLS data when performing most of these tasks requires addressing a previous and fundamental step, namely the registration of the sonar images. Although registration is a broadly studied field in other modalities, notably the optical one (Zitova & Flusser, 2003), it is still a premature field with regard to sonar data. The particularities of FLS imagery, such as low resolution, low signal-to-noise ratio (SNR), and intensity alterations due to viewpoint changes, pose serious challenges to the feature-based registration techniques that have proved very effective at aligning optical images.

The need to find a registration technique suited to FLS images has led researchers to investigate the problem through different approaches. Most of the existing work adopts a feature-based pipeline where feature detection is performed either by using detectors at pixel scale (Kim, Intrator, & Neretti, 2004; Negahdaripour, Firoozfam, & Sabzmejdani, 2005) or by looking for more stable features extracted at region level (Aykin & Negahdaripour, 2012; Johannsson et al., 2010). Leaving aside the ability of these methods to cope with the noise and artifacts in sonar images, it is clear that they require the presence of prominent features in the environment that can be unequivocally matched. In general, the fewer the features, the lower the possibility of establishing successful registrations, thus impacting the effectiveness of subsequent processing. Moreover, the difficulties in accurately extracting and matching

stable features are exacerbated when dealing with spatially or temporally distant sonar images found in loop closure situations. This is a key issue since the registration of revisited locations is crucial to bound the error accumulated over time and achieve global consistency in mosaicing or motion estimation applications.

In this work, we propose using a Fourier-based technique to perform 2D FLS registration (Hurtós, Cufi, Petillot, & Salvi, 2012). Instead of making use of sparse feature information, we propose to use a global method that takes into account the whole image content, thus contributing to the minimization of ambiguities in the registration. The method is, by design, robust to noise and inhomogeneous insonification, and it handles well the challenging nature of sonar images, achieving a high degree of success in registering not only consecutive frames but also revisited frames in loop closure situations. Without requiring the extraction of explicit features, the method is independent of the type and number of features present in the environment and can be robustly applied to a wide variety of applications ranging from surveying natural terrain to inspecting manmade scenarios.

Our purpose is to exploit this registration methodology as a robust way to map underwater environments under low-visibility conditions by using autonomous or remotely operated vehicles (ROVs). Our first focus is the generation of 2D acoustic mosaics of underwater areas of interest. Possibly due to the recent introduction of high-resolution FLS devices, the specific problem of FLS mosaicing has only been tackled by a few researchers (Kim et al., 2005; Negahdaripour et al., 2005, 2011) together with some related work dealing specifically with FLS image registration (Aykin & Negahdaripour, 2012; Johannsson et al., 2010). Nevertheless, the existing mosaicing approaches have shown very limited results in terms of scale and complexity, as most of the reported mosaics are restricted to only a few frames gathered in a single straight trackline while imaging feature-rich scenarios. Here we propose a complete mosaicing pipeline that enables the creation of consistent mosaics extending along various vehicle tracklines undergoing both translational and rotational 2D motions, and applicable in a wide variety of environments, including those with a scarcity of features.

Toward that end, we utilize the proposed registration technique to compute pairwise constraints between sonar frames that are later embedded in a pose-based graph formulation to enforce global alignment. This enables the rendering of consistent 2D acoustic mosaics of high detail that not only offer a global overview of the surveyed area, but provide a significant improvement of the SNR and resolution with respect to the individual images.

By following the same scheme, the registration technique can be used to extract 2D vehicle motion estimates from the sonar imagery. Although we present herein an offline framework in which the estimation of the trajectory is

computed *a posteriori*, we believe that the proposed method is amenable to being integrated in an online SLAM framework (Kaess, Ranganathan, & Dellaert, 2008) to perform sonar-aided navigation.

The remainder of this paper is organized as follows. Section 2 provides a background on FLS imaging, analyzing the geometry model and the challenges encountered when working with sonar images. In Section 3, the proposed registration method is presented and its performance is analyzed against a state-of-the-art technique. Section 4 covers the global alignment stage performed by means of a pose-based graph optimization. Section 5 deals with the insights of the sonar mosaic composition. Experiments with real datasets including relevant field applications such as ship hull inspection and harbor mapping are described in Section 6 together with the corresponding results. The final section concludes the paper and points out future work prospects.

## 2. BACKGROUND ON FORWARD-LOOKING SONAR IMAGING

To address the registration of FLS data, it is necessary to understand the image formation process and find a suitable model to describe the imaging geometry of the sonar. The following is a description of the mode of operation of FLS, a review of the FLS geometry models used in the related state of the art, and a discussion on our model choice together with its limitations. We also provide a summary of the main challenges to be faced when dealing with FLS imagery to better understand how they effect the registration process.

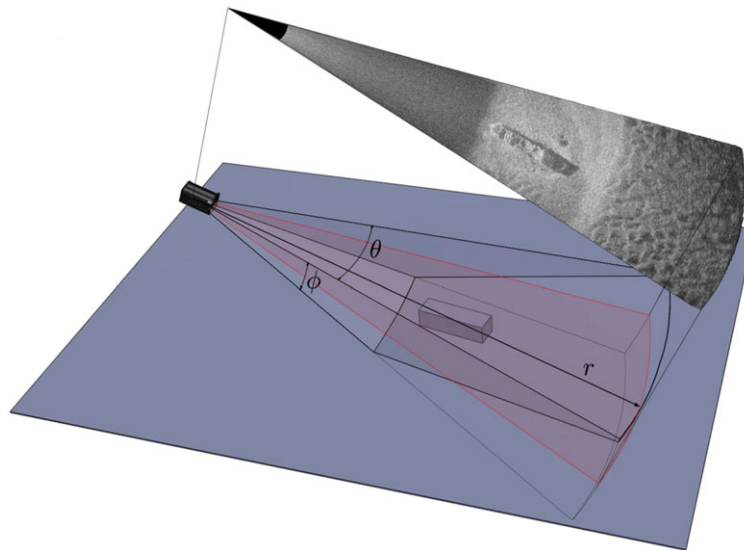
### 2.1. FLS Operation

2D FLSs, sometimes also referred to as acoustic cameras, are a novel category of sonars that provide high-definition acoustic imagery at a fast refresh rate. Although the specifications regarding operating frequency, acoustic beamwidth, frame rate, and the internal beamforming technology depend on the specific sonar model and manufacturer, the principle of operation is the same for all. First, the sonar insonifies the scene with an acoustic wave, spanning its field of view (FoV) in the azimuth ( $\theta$ ) and elevation ( $\phi$ ) directions, and then the acoustic return is sampled by an array of transducers as a function of range and bearing (Figure 1). Because of the sonar construction, it is not possible to disambiguate the elevation angle of the acoustic return originating at a particular range and bearing. In other words, the reflected echo could have originated anywhere along the corresponding elevation arc. Hence, the 3D information is lost in the projection into a 2D image.

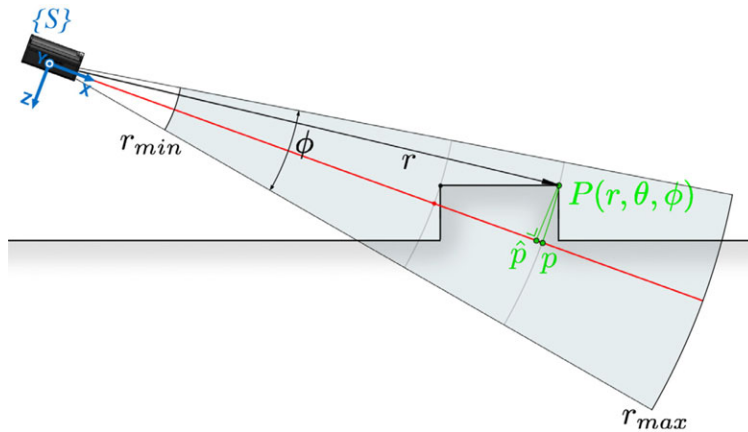
### 2.2. Imaging Geometry Model

According to the described principle of operation, a 3D point  $\mathbf{P}$  with spherical coordinates  $(r, \theta, \phi)$  can be defined in the sensor frame  $\{S\}$  by the following Cartesian coordinates:

$$\mathbf{P} = \begin{bmatrix} X_s \\ Y_s \\ Z_s \end{bmatrix} = \begin{bmatrix} r \cos \theta \cos \phi \\ r \sin \theta \cos \phi \\ r \sin \phi \end{bmatrix}. \quad (1)$$



**Figure 1.** The sonar emits an acoustic wave spanning its beam width in the azimuth ( $\theta$ ) and elevation ( $\phi$ ) directions. Returned sound energy is sampled as a function of  $(r, \theta)$  and can be interpreted as the mapping of 3D points onto the zero elevation plane (shown in red).



**Figure 2.** The sonar projection geometry maps a 3D point  $P(r, \theta, \phi)$  into a point  $p$  on the image plane along the arc defined by the elevation angle. Considering an orthographic approximation, the point  $P$  is mapped into  $\hat{p}$ , which is equivalently to considering that all scene points rest on the plane  $XY_s$  (in red).

This 3D point  $P$  is projected in a point  $\mathbf{p} = (x_s, y_s)$  on the image plane ( $XY_s$ ) following a nonlinear model:

$$\mathbf{p} = \begin{bmatrix} x_s \\ y_s \end{bmatrix} = \begin{bmatrix} r \cos \theta \\ r \sin \theta \end{bmatrix} = \frac{1}{\cos \phi} \begin{bmatrix} X_s \\ Y_s \end{bmatrix}. \quad (2)$$

As can be seen in Eq. (2), the projection is introduced as a function of the elevation angle. Depending on the treatment of this projection, we can distinguish two different ways of approaching FLS geometry.

Given the narrow elevation angle that typically characterizes FLS devices (around  $6^\circ$ – $10^\circ$ ), the nonlinear component defined by  $\phi$  is tightly bound. Approximating this narrow elevation to the limit (i.e., considering only the zero-elevation plane), we end up with a linear model in which the sonar can be seen as an orthographic camera (Walter, 2008). Hence, the projection  $\mathbf{p}$  of a 3D point  $P$  is approximated by the orthogonal projection  $\hat{\mathbf{p}}$  as shown in Figure 2.

Analogously to the parallax problem in optical views, this approximation holds as long as the scene’s relief in the elevation direction is negligible compared to the range, as the error introduced by the projection approximation is a function of the distance in the  $XY_s$  plane and the vertical distance to the point (Johannsson et al., 2010). The imaging geometry under a typical operation scenario falls within this consideration since the sonar device is normally tilted to a small grazing angle to cover a large portion of the scene. On the other hand, the projection preserves the change in azimuth angles, i.e., if the sonar rotates with respect to its vertical axis, the projection on the image rotates by the same angle. Rotation around pitch, usually not present or controlled by a tilt unit, affects the limits of the imaged area and its reflected intensities but does not introduce a change in the projection of the points. Changes in roll would affect the  $y$ -axis of the projections, but we consider it negligible

due to the usual stability of underwater vehicles in this degree of freedom.

Therefore, by using this model, a point in the space represented by  $\mathbf{p}$  and  $\mathbf{p}'$  in two different images can be related through a global affine homography  $\mathbf{H}$ . This homography describes the 2D motion from one position to the next in terms of a 2D rigid transformation comprising the  $x$  and  $y$  translations ( $t_x, t_y$ ) and the plane rotation ( $\theta$ ):

$$\mathbf{p}' = \mathbf{H}\mathbf{p} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & t_x \\ \sin(\theta) & \cos(\theta) & t_y \\ 0 & 0 & 1 \end{bmatrix} \mathbf{p}. \quad (3)$$

Other approaches (Aykin & Negahdaripour, 2012; Sekkati & Negahdaripour, 2007) work on the exact model, without considering the narrow elevation approximation. Then, the homography  $\mathbf{H}$  relating two image points  $\mathbf{p}$  and  $\mathbf{p}'$  becomes an affine homography whose elements vary across the image depending on the range and the unknown elevation angles (Negahdaripour, 2012b):

$$\mathbf{p}' = \mathbf{H}\mathbf{p} = \begin{bmatrix} \alpha q_{11} & \alpha q_{12} & \beta q_{13} \\ \alpha q_{21} & \alpha q_{22} & \beta q_{23} \\ 0 & 0 & 1 \end{bmatrix} \mathbf{p}, \quad (4)$$

where  $\alpha = \cos \phi / \cos \phi'$ ,  $\beta = r \sin \phi / \cos \phi'$ , and  $q_{ij}$  denotes the  $i, j$  components of a matrix  $\mathbf{Q} = \mathbf{R} - \mathbf{t}\mathbf{n}^T$  that is the rigid-body motion transformation for features lying on a plane with normal  $\mathbf{n}$ . Hence, the imaging model is a nonuniform function of the image coordinates and the surface normal of the assumed underlying plane, with  $\mathbf{H}$  encoding all the information about the 3D sonar motion and surface parameters. The differential version of this model, dealing with rotational and translational velocity components (Negahdaripour, 2012a), has also been used in the context

of 3D sonar motion estimation (Aykin & Negahdaripour, 2013). In theory, it allows handling all six degrees of freedom (DoF) of the sonar motion, while in practice the pitch-and-roll motion components are not estimated due to sensitivity to various sources of error and noise in the sonar data (Negahdaripour, 2012a).

However, using these models requires knowledge of the elevation angles at every image location, which is not provided by the sonar. Negahdaripour has shown that an elevation map of the imaged plane can be determined from its surface normal. An estimation of the surface normal can be computed from the sonar range settings and the imaging configuration relative to the scene (Negahdaripour, 2012a). Moreover, the elevation map can be defined with higher accuracy by incorporating the elevation angles of prominent features. Aykin and Negahdaripour make use of object-shadow pairs extracted from detected blob regions to estimate the elevation angle of 3D features.

The advantages these models offer when compared to the simplified 2D model are the estimation of the sonar motion in the vertical direction ( $z$ ) and a more accurate registration as a result of accounting for the elevation angles at each location. On the one hand, estimation of the sonar motion in the  $z$  direction is not required for 2D mapping purposes. If the sonar motion were to be estimated, the translations in the vertical direction could be reliably obtained from pressure sensors. It is thus sufficient to estimate the  $x, y$  translations and yaw rotation, which are the measures affected by drift and bias, respectively. On the other hand, the incorporation of the elevation angles in the registration process reduces the errors introduced by the orthographic approximation and has proved to enhance the local image alignment (Negahdaripour, 2012a). Nevertheless, this is subject to the ability of robustly estimating the elevation angles on the imaged surface, which may not be a trivial procedure depending on the imaging configuration or the type of features present in the environment.

Therefore, in this work, we chose to adopt the simplified 2D model. Although it is an approximation, it is suitable to describe the image formation process and set the basis for the subsequent registration process. Using a model of only 3 DoF allows us to consider global-area registration techniques that resolve only fixed transformations applied to the entire image.

It is worth emphasizing that the main limitation of this model, namely the assumption of the imaged scene being nearly planar, can be relaxed thanks to the range length of the FLS devices, which can vary from 10 to 50 m depending on the sonar. These ranges offer the flexibility of adopting a more appropriate imaging configuration so that the assumption of the projections lying on a plane becomes more realistic, i.e., imaging from a farther distance or at a narrower grazing angle while still achieving an acceptable resolution. Note that in the optical case, this flexibility is constrained by the light attenuation and the short visibility

ranges of underwater cameras. Besides, the use of a pan and tilt unit together with sensors that can provide an estimation of the underlying plane (e.g., profiling sonars or multibeam systems) can be considered to accommodate the imaging configuration so as to match the horizontal assumption as closely as possible.

### 2.3. Challenges in FLS Imagery

Acoustic images offer the ability to see through turbid environments at the expense of dealing with a much more challenging type of data. There are some particularities closely related to the nature of sonar image formation that increase the difficulty of their registration, especially when compared to optical images.

- **Low resolution:**  
Although they are considered high-definition sonars, 2D-FLS image resolution is far from the resolution of today's standard cameras that make use of 2D array sensors with millions of pixels. For instance, the ARIS sonar (Sound Metrics ARIS, 2013) samples the acoustic returns with an array of 128 transducers with a  $0.3^\circ$  beamwidth. BlueView P900-130 (BlueView Technologies Inc., 2013) has 768 beams with  $1^\circ$  of beamwidth each. Moreover, as a consequence of the sensor's polar nature, measurement sparseness increases with the range when represented in a Cartesian space. This results in a nonuniform resolution that degrades the image's visual appearance.
- **Low signal-to-noise ratio:**  
As with other coherent imaging systems such as radar or ultrasound imaging, 2D FLS suffers from low SNR. This is mainly due to the presence of speckle noise introduced by the mutual interference of the sampled acoustic returns.
- **Inhomogeneous insonification:**  
FLS is commonly affected by inhomogeneous intensity patterns due to differing sensitivity of the lens or transducers according to their position in the sonar's field of view (Negahdaripour et al., 2005). These intensity patterns can affect the registration, causing the images to align on them instead of the real image content. However, this can be corrected by means of a preprocessing step that estimates the inhomogeneous intensity pattern from the averaging of a sufficient number of images.
- **Changes in viewpoint:**  
Intensity variations due to a change in the sonar's viewpoint are inherent in the image formation process. Imaging the same scene from two different vantage points can cause the movement of shadows in the images, occlusions, and, in general, significant alterations in the visual appearance of the content that complicate the registration process. To minimize these effects, it is preferable to image the area always from the same sonar point of view, though this might not be always feasible. Hence, it is desirable that the registration algorithm can

cope with alterations caused by substantial viewpoint changes.

- Other artifacts:

Under some circumstances, spurious content can appear in the sonar images causing ambiguity in the registration: reverberation artifacts, acoustic returns from the water surface, or cross-talk between beams that generates multiple replicas of a target. However, these artifacts can generally be minimized by adopting a proper configuration and imaging setup.

### 3. PAIRWISE REGISTRATIONS OF FLS

The computer vision community has proposed numerous registration methods over the past few decades (Zitova & Flusser, 2003). Among the most popular are the feature-based methods that rely on the detection of a limited set of well-localized individually distinguishable points (Tuytelaars & Mikolajczyk, 2008). The traditional pipeline for feature-based registration of images consists, first, of the detection of local features followed by a feature extraction process. The extraction is usually performed by computing descriptors, i.e., a compact representation of the neighborhood of a feature. Afterward, there is a matching step in which the point-to-point correspondences from the two images are established, and, finally, this information is used to estimate the transformation that relates one image to the other, usually by taking into account an outlier rejection scheme such as RANSAC (Fischler & Bolles, 1981).

Some of these feature-based approaches have been applied to the registration of FLS images. In general, reported results come from small and feature-rich datasets, and registrations are performed only between consecutive frames. In Negahdaripour et al. (2005), a few image pairs from a DIDSON sonar are registered using a Harris corner detector and matched by searching over small local windows. Similarly, in Kim et al. (2005), Harris features extracted at the third and fourth level of a Gaussian pyramid scale are matched with cross-correlation and used in a mosaicing algorithm. Each frame is registered sequentially with a window of neighboring frames, and results show only the registration from translational sonar displacements. Negahdaripour *et al.* highlight the complexities of mosaicing benthic habitats with FLS images (Negahdaripour et al., 2011) and show the difficulty of registering DIDSON frames from a natural environment by using the popular SIFT detector (Lowe, 2004). Results report a very low percentage of inliers in the detection step (about 8%), and only small displacements could be effectively matched.

In general, due to the inherent characteristics of sonar data, pixel-level features extracted in sonar images suffer from low repeatability rates (Hurtós, Nagappa, Cuff, Petillot, & Salvi, 2013b). Consequently, they lack stability and are prone to yielding wrong transformation estimations.

This fact has not gone unnoticed by other researchers, who have proposed alternatives involving features at region level rather than at pixel scale. Johannsson et al. (2010) proposed the extraction of features in local regions located on sharp transitions (i.e., changes from strong to low signal returns as in the boundaries of object-shadow transitions). The sonar images are first smoothed with a median filter, then their gradients are computed, and points exceeding a given threshold are finally clustered in features. These features are presumably more stable than those computed at pixel level. Feature alignment is formulated as an optimization problem based on the normal distribution transform (NDT) algorithm (Biber & Straßer, 2003). The NDT adjusts the clustered regions in grid cells, removing the need to get exact correspondences between points, thus allowing for possible intensity variations. However, the registration accuracy becomes strongly dependent on the selected grid resolution.

A similar approach has been recently presented by Aykin & Negahdaripour (2012). Instead of thresholding on the gradient domain, the highest intensity values in the images (assumed to be objects or structures on the ground surface) are clustered in blob features. As an alternative to the NDT algorithm, Aykin and Negahdaripour propose the use of an adaptive scheme in which a Gaussian distribution is fitted to each blob feature. Afterward, an optimization is formulated to seek the motion that best fits the blob projections from one Gaussian map to the other.

Taking it a step further, it seems natural to explore area-based methods that, instead of using sparse feature information, make use of the entire image content. By incorporating more information in the registration process, we are able to handle more changes in the visual appearance of the image and minimize the ambiguities in the registration. The common shortcoming of area-based techniques is that they cannot handle complex transformations, being limited to the estimation of similarity transforms. However, and according to the simplified FLS geometry model that we adopted, the registration of two FLS images falls inside its scope of applicability, thus turning the area-based methods into a candidate solution for FLS image alignment. From all the different area-based approaches, we propose the use of Fourier-based techniques. The particularities of these methods suggest that they might be appropriate for the registration of FLS imagery since, by design, they offer robustness to noise, illumination changes, and occlusions (Foroosh, Zerubia, & Berthod, 2002). In this section, we will describe insights into the proposed Fourier-based registration technique, and then we will compare its performance with a state-of-the-art region-based methodology.

#### 3.1. Fourier-based Registrations for FLS

Fourier-based methods, in particular the phase correlation algorithm (De Castro & Morandi, 1987; Reddy & Chatterji,

1996), have been successfully employed in several image processing tasks, such as image registration, pattern recognition, motion compensation, and video coding, to name a few. These techniques allow registrations up to similarity transformations with a high computational efficiency due to the implementation of the fast Fourier transform (FFT) algorithm. In a similar problem to the one we tackle in this work, phase correlation has been applied to register underwater optical images in order to build photomosaics (Bülow, Birk, & Unnithan, 2009; Eustice, Pizarro, Singh, & Howland, 2002). However, when dealing with video images, feature-based methods are generally more popular since their high resolution and SNR allow us to extract stable features easily and estimate more general transformations such as projective homographies.

On the other hand, the literature regarding the application of Fourier-based methods on sonar imagery is not extensive. Some authors have pointed out the phase correlation method as potentially useful in the registration of side-scan sonar images (Chailloux, 2005; Vandrish, Vardy, Walker, & Dobre, 2011), while other researchers employed it in the registration of 2D and 3D sonar range scans (Bülow & Birk, 2011; Bülow, Pflingsthorn, & Birk, 2010).

According to the Fourier shift property, a shift between two functions (e.g., images) is transformed in the Fourier domain into a linear phase shift.

Let  $f(x, y)$  and  $g(x, y)$  be two images related by a 2D shift  $(t_x, t_y)$ , namely

$$f(x, y) = g(x - t_x, y - t_y). \quad (5)$$

Then their 2D Fourier transforms, denoted by  $F(u, v)$  and  $G(u, v)$ , are related via

$$F(u, v) = G(u, v)e^{-i(ut_x + vt_y)}. \quad (6)$$

Their normalized cross power spectrum is given by

$$C(u, v) = \frac{F(u, v)G^*(u, v)}{|F(u, v)G^*(u, v)|} = e^{-i(ut_x + vt_y)}, \quad (7)$$

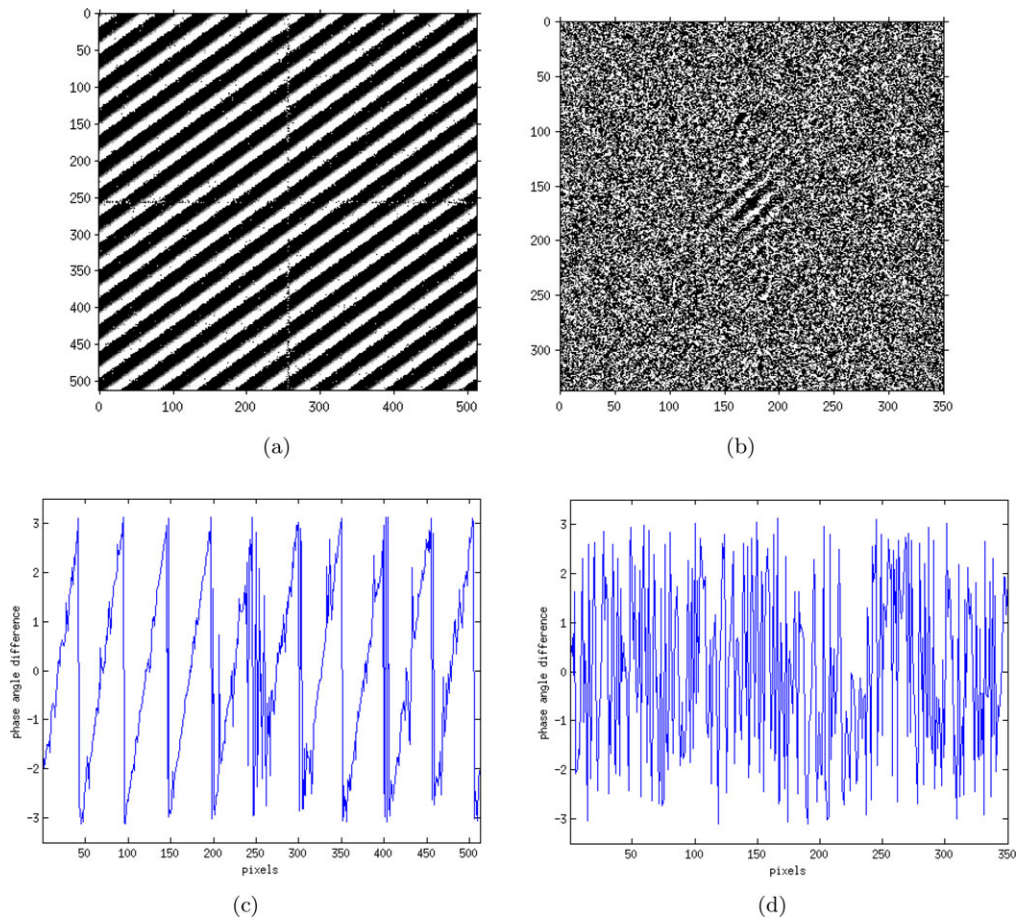
where  $G^*$  denotes the complex conjugate of  $G$ . The normalizing denominator in this equation is equivalent to a prewhitening of the signals, making the phase correlation method inherently robust to noise that is correlated with the images, such as uniform variations of illumination or offsets in average intensity (Feroosh et al., 2002). The most common way to solve Eq. (7) for  $(t_x, t_y)$  is to apply the inverse Fourier transform to  $C(u, v)$ , obtaining the phase correlation matrix (PCM). In the ideal case, this matrix corresponds to a 2D impulse (Dirac function) centered on  $(t_x, t_y)$  that directly leads to the identification of the integer displacements. In the presence of noise or other image perturbations, the Dirac pulse of the phase correlation matrix deteriorates, but as long as it contains a dominant peak, the offsets can be retrieved. Moreover, after determining the maximum correlation peak with integer accuracy, subpixel displacements can be estimated (Feroosh et al., 2002; Ren, Jiang, & Vlachos, 2010).

A different group of approaches try to recover the offsets in Eq. (7) by working only in the frequency domain. The shifts are then computed as the slopes of a plane fitted to the phase difference data (Hoge, 2003). Balci & Feroosh (2006) have shown that the phase difference matrix of two images is a 2D sawtooth signal whose cycles determine the shift parameters. Therefore, it is only necessary to robustly count the number of cycles along each frequency axis to retrieve the translational offsets. While there are multiple publications reporting successful results with optical images (Balci & Feroosh, 2006; Hoge, 2003), the implementation of the Balci and Feroosh method does not seem feasible with FLS images. Figure 3 shows an example of the phase difference matrices obtained from two optical images [Figure 3(a)] and two FLS images [Figure 3(b)]. While it is possible to compute reliably the length of a sawtooth cycle in the optical case, the cycles are hardly distinguishable in the FLS example. Even after attempting filtering operations, the robust estimation of the offsets from the phase difference cycles is unfeasible. In our experience, working directly in the frequency domain offers a much higher sensitivity to noise compared to computing the inverse transform of the cross power spectrum and finding the peak in the spatial domain, the reason for which we developed in the standard approach.

There are a number of factors that may introduce arbitrary peaks in the PCM, thus reducing the ability to detect a clear dominant peak. The challenges described in Section 2.3, such as low SNR or intensity alterations due to different vantage points, are likely to give rise to multiple local maxima in the PCM and reduce the amplitude of the true registration peak. A similar effect occurs as a consequence of the content of nonoverlapping image areas or due to errors introduced by the approximations of the adopted geometry model.

It is common practice to apply some filtering operations to the image's spectra in order to attenuate unwanted frequencies that can lead to a noisy phase correlation matrix (Reddy & Chatterji, 1996). However, determining these filters can be critical as there is a risk of attenuating not only the unwanted components but also the discriminating phase components. We seek to capture both low-frequency characteristics, such as the change of reflectivity from a sandy area to vegetation, and high-frequency components that arise from object edges or protruding seabed features. Therefore, we decided not to apply any filtering prior to the computation of the phase correlation matrix. Once back in the spatial domain, a small smoothing filter is applied to reduce the noise and enhance the robustness of the peak detection.

Additionally, there are some factors not linked to the image nature itself that can lead to failure in detecting the correlation peak if not handled properly. The most critical are the so-called edge effects. The phase correlation theory described previously holds for periodic signals and



**Figure 3.** Example of the Balci and Foroosh method. (a) Phase difference matrix corresponding to a pair of shifted optical images. (b) Phase difference matrix corresponding to a pair of shifted sonar images. (c),(d) One row of (a) and (b), respectively. Notice the difficulty of detecting the cycles and, therefore, the shifts in the sonar case.

continuous Fourier transforms. In the discrete case, the FFT is used to approach the infinite Fourier transform, imposing a cyclic repetition of finite-length images. The abrupt transitions generated between the edges when the images are tiled result in high-frequency components appearing in the Fourier spectrum, which may alter the subsequent computation of the phase correlation matrix. In a similar manner, the fan-shaped boundaries of the FLS images in Cartesian coordinates introduce high-frequency components that do not depend on the image content. That causes a strong false peak around the origin of the phase correlation matrix that can hide the location of the true peak. To minimize these effects, it is typical to perform a windowing operation before the FFT computation. In our case, a mask that tapers the boundaries of the FLS images in Cartesian coordinates is applied to the images prior to the FFT computation.

Up to this point, we have referred to the estimation of linear shifts from the sonar images. However, the recovery

of rotations must also be addressed. The inherent nature of sonar images suggests that mapping an area while maintaining the same viewpoint orientation increases the number of successful registrations between sonar frames. This way, a lawn mower pattern in which the transition from track to track is performed by sway displacement instead of rotation would be a good mapping strategy. However, this approach might not always be feasible, either because the vehicle does not allow for the sway degree of freedom, or simply because the area to cover does not follow a rectangular layout and requires some orientation changes in order to be efficiently covered. Moreover, if we think not only about autonomous surveys but inspections carried out with remotely operated vehicles as well, the pilot will most likely undertake a great number of rotational movements. Hence, it is important to find a robust solution to estimate the rotation between pairs of FLS images so as to enable the use of sonar mapping in more diverse situations and environments.



In a previous work (Hurtós, Cufi, & Salvi, 2014), we evaluated the performance of several global-area methods for rotation estimation on real FLS images. The general outcome is summarized here, and the reader is referred to Hurtós et al. (2014) for further details.

One of the most popular methods dealing with the estimation of rotational alignments is based on the polar magnitude of the Fourier transform, often referred to as the Fourier-Mellin transform (Chen, Defrise, & Deconinck, 1994; Reddy & Chatterji, 1996). According to the Fourier shift property, translational displacements affect only the phase spectrum, while the magnitude is invariant to them. Therefore, since a rotation is mapped as a linear shift in the angular direction of the polar domain, it can be recovered in a manner invariant to the translation by using the polar magnitude of the Fourier transform. The rotation estimation problem is then converted to a shift estimation where the input images are the polar representations of the Fourier transform magnitudes. This shift estimation can be solved by standard phase correlation, and leads to two possible solutions ( $\theta$  and  $\theta + \pi$ ) that can be disambiguated by trying to solve for the translation in each case and keeping the one that leads to the highest correlation peak. This technique, widely popular in optical images (Bülow et al., 2009; Schwertfeger, Bülow, & Birk, 2010), is not as robust in the case of FLS imagery. Applying phase correlation to the polar representation of the magnitude spectrum leads to erroneous results since it has a low structural nature (which is even lower in the case of sonar modality) and suffers from inaccuracies introduced by the interpolation process to the polar domain.

Likewise, other popular techniques are deemed unfeasible when applied to FLS imagery, either because they are targeted for images with high resolution and high SNR (Keller, Shkolnisky, & Averbuch, 2005; Lucchese & Cortelazzo, 2000) or because they become expensive time-wise when aiming for a certain level of accuracy and robustness (Costello, 2008; Li et al., 2007).

In view of all this, we considered the option of estimating the rotation as a shift displacement directly on the polar images rather than working with the magnitude of its Fourier transformation. In this way, the estimation is performed on the raw data delivered by the sensor, thus avoiding any interpolation or the need to work with representations of the Fourier transform. However, when working with the polar images, rotation is not decoupled from translational displacements, and shifts in Cartesian space create distortions in the polar domain. If the translational displacements are relatively small compared to the image's size in each direction, the induced distortions in the polar image still allow for the recovery of the rotation by computing the shift in the angular direction. The high frame rate of FLS devices facilitates large overlaps and therefore small translations between consecutive and near-consecutive frames, thus not affecting the rotation estimation under this scheme.

Moreover, there are cases in which rotations are not combined with translations (the vehicle stops, rotates, and then continues), yielding a pure translation in the polar domain. The major drawback is then in loop-closing situations, when attempting to match temporally distant frames that present significant shifts. In these cases, the proposed strategy for rotation estimation is prone to introducing inaccuracies in the estimated angle. This, in turn, affects the number of encountered loop closures, as the loop closures that involve more overlap (i.e., smaller translations) and smaller orientation changes are more likely to be successfully registered. Nevertheless, as will be seen in Section 4.2, these inaccurate estimations can be identified with the help of a measure that quantifies the uncertainty of the registration, and therefore we can prevent them from having a negative effect in subsequent processing.

It is important to note that, by construction, this method does not allow for the estimation of an angle difference higher than the FoV of the sonar. This limit becomes even more restricted if we take into account that a minimum overlap is required in order to establish the correlation. For instance, in cases of pure rotation and aiming for a minimum overlap of 50%, the limits of the rotations that can be estimated are within  $[-\frac{\text{FoV}}{2} : \frac{\text{FoV}}{2}]$  degrees. If translations are also involved, the overlap will decrease, thus reducing even more the possibilities of estimating the rotation correctly. This is a fairly strong restriction, especially in sonars with narrow fields of view. However, due to the high frame rate of FLS devices, sequential and near-sequential images typically undergo small rotations easily falling inside those limits, and, therefore, guaranteeing the establishment of local constraints under the presence of rotations. In loop closure situations, it is more difficult to conform to that restriction. However, it is to our advantage to choose a mapping strategy that allows revisiting locations with orientations comprised within these limits. Furthermore, if information of the path topology is known in advance, we can determine beforehand if the images belong to tracks with reciprocal headings. When this is the case, the polar frames are flipped before performing the phase correlation, thus leading to the estimation of rotations comprised within  $[-\frac{\text{FoV}}{2} + 180 : 180 + \frac{\text{FoV}}{2}]$  degrees.

Despite these limitations, this rotation estimation method outperforms the rest of the mentioned techniques in terms of robustness and accuracy (Hurtós et al., 2014) and is the one employed in our registration pipeline. In the next section, its performance will be compared to the traditional Fourier-Mellin approach that estimates rotation by using phase correlation on the polar magnitude of the image's FTs.

The flowchart of Figure 4 outlines the general procedure to register two sonar images. The sonar frames in polar coordinates ( $f_p, g_p$ ) are first masked by a cosine window to avoid edge effects arising from the image's boundaries. Using the transforms of these images as

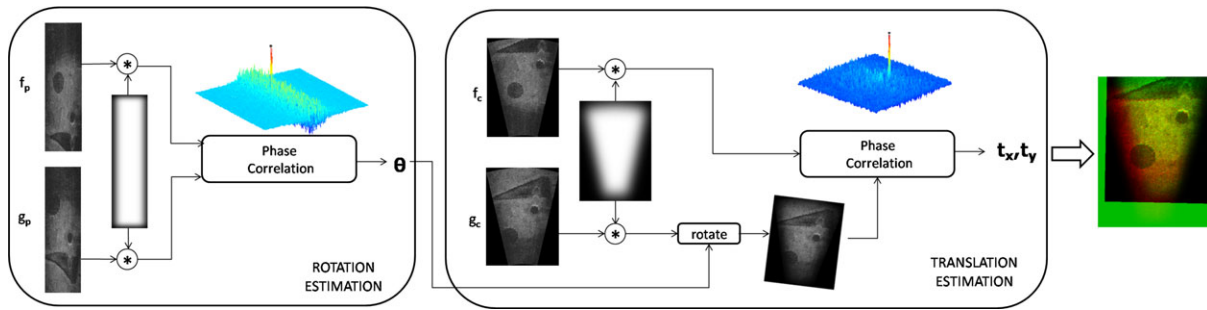


Figure 4. Overall registration pipeline.

input, phase correlation is applied following Eq. (7). The shift detected in the  $x$  direction provides an estimate of the rotation angle  $\theta$  between the images in Cartesian coordinates  $(f_c, g_c)$ . After masking both images with the corresponding Cartesian mask, phase correlation is performed between  $f_c$  and the rotation-compensated version of  $g_c$  to finally obtain the translations in the  $x$  and  $y$  directions that align them.

### 3.2. Comparison with Region-based Registration

In this section, the performance of the proposed Fourier-based method is compared against a state-of-the-art FLS registration technique. Comparison with feature-based methods at pixel level has been discarded, since, as explained in Section 3, its poor performance with FLS images is well-known and has been reported (Hurtós et al., 2013b; Negahdaripour et al., 2011; Walter, 2008). From the two existing region-based methods for FLS alignment (Aykin & Negahdaripour, 2012; Johannsson et al., 2010), we select the method of Johannsson *et al.* This selection is motivated by the geometry model under consideration: Aykin and Negahdaripour assume a 3D sonar motion model that incorporates the unknown elevation angles in the registration process, while Johannsson *et al.* work with the same 2D simplified model that we adopt.

We have implemented the technique of Johannsson *et al.* following their steps as described in Hover *et al.* (2012). The feature extraction process applies median smoothing on the image followed by gradient computation. The gradient is computed as the difference between a value and the mean of its  $n$  previous values along its azimuthal line. Then, a fixed fraction of points with negative gradient are segmented and clustered in features. The registration of these features is performed using the NDT algorithm with four overlapping grids shifted half a cell. The NDT implementation of the point cloud library (PCL) (Point Cloud Library, 2013) has been used for this step. Following the same procedure as the authors, the NDT optimization is performed several times with different initialization points.

To compare both methods, we have used three datasets in which the ground truth is available. These datasets allow us to test the registration under different conditions, including different sonar models and different motion types. The first dataset is comprised of 944 sonar frames gathered with an ARIS sonar (Sound Metrics ARIS, 2013) inside a harbor. The FLS was mounted on a pole together with a GPS and attached to a boat. The sequence follows a straight transect with mainly translational displacements in the  $x$  direction. According to the sonar's configuration, the range resolution is 8 mm/pixel and angular resolution is  $0.2^\circ$ . The second dataset consists of 1,176 sonar frames gathered with a DIDSON (Sound Metrics DIDSON, 2013) in a dock environment. The sonar performed a  $360^\circ$  scan with steps of  $0.3^\circ$  mounted on a tripod. These rotational increments correspond to the sonar's angular resolution, while the range resolution is approximately 1.9 cm/pixel. The last dataset was gathered with a BlueView P900-130 (BlueView Technologies Inc., 2013) in a harbor environment with an Autonomous Surface Catamaran (see further details of the dataset in Section 6.3), performing both rotational and translational motions. Range and angular resolution are 6 cm/pixel and  $0.3^\circ$ , respectively. Therefore, the estimated translations and rotations will be compared using as ground truth the GPS locations in the first and third dataset and the fixed mechanical tripod step in the second dataset. It is worth noting that we make use of a high-precision RTK GPS that also delivers an accurate heading by employing a setup with two antennas. Moreover, a large number of registration results are averaged in each case. In this way, we consider that the effect of any possible GPS errors over the reported mean errors is negligible.

Before carrying out the comparison between the region-based and the Fourier-based registrations, we employ the described datasets to compare the proposed rotation estimation method with the traditional Fourier-Mellin approach. Table I presents the mean and maximum rotation errors with respect to the ground truth when estimating the rotation between consecutive frames for the different datasets. The same experiment has been repeated by choosing more distant frames this time, overlapping about 60% (Table II).

**Table I.** Comparison experiments between rotation estimation methods when registering consecutive frames.

	Fourier-Mellin		Directly on polar images	
	Mean error (deg)	Max error (deg)	Mean error (deg)	Max error (deg)
<b>Dataset 1</b>	0.92	1.40	0.51	0.61
<b>Dataset 2</b>	0.64	0.83	0.03	0.42
<b>Dataset 3</b>	1.08	7.51	0.54	7.60

**Table II.** Comparison experiments between rotation estimation methods when registering distant frames.

	Fourier-Mellin		Directly on polar images	
	Mean error (deg)	Max error (deg)	Mean error (deg)	Max error (deg)
<b>Dataset 1</b>	1.46	4.30	1.15	3.91
<b>Dataset 2</b>	2.13	4.07	0.09	5.52
<b>Dataset 3</b>	3.02	21.7	1.72	29.5

In all cases, even when estimating the rotation of distant frames, the estimation through direct phase correlation on the polar images leads to better accuracy than performing the estimation on the polar magnitude of the image’s FFTs. The differences are especially significant for the second dataset, in which clearly the proposed method is highly accurate due to the presence of pure rotations. Nevertheless, the mean errors in the other cases are also lower for the proposed method, thus testifying to the fact that the noise and the low structural nature of sonar images makes the robust correlation of the polar FFT magnitudes difficult.

Figure 5 shows three images illustrative of each dataset together with examples of extracted features. For each dataset, two different tests have been performed. The first consists of registering each sonar frame with its consecutive in the sequence. The second test aims to compare the performance of the methods when dealing with spatially and temporally distant images. Given that not all available datasets comprise trajectories with loop closures, the test attempts the registration of a frame with a neighbor frame in the sequence. The interval between frames is chosen for each dataset in order to reduce the overlap to approximately 60%. Although the changes induced in the images may not be as severe as in an actual loop closure situation, they are sufficient to evaluate the trends of the methods when dealing with distant images.

Before applying the method of Johannsson *et al.* to each sequence, several tests have been performed to tune its parameters according to the dataset’s characteristics and the image content. Therefore, the value  $n$ , the gradient thresh-

old, and the number of extracted points have been adjusted to achieve a good balance of extracted features. Likewise, the grid size of the NDT algorithm has been modified appropriately.

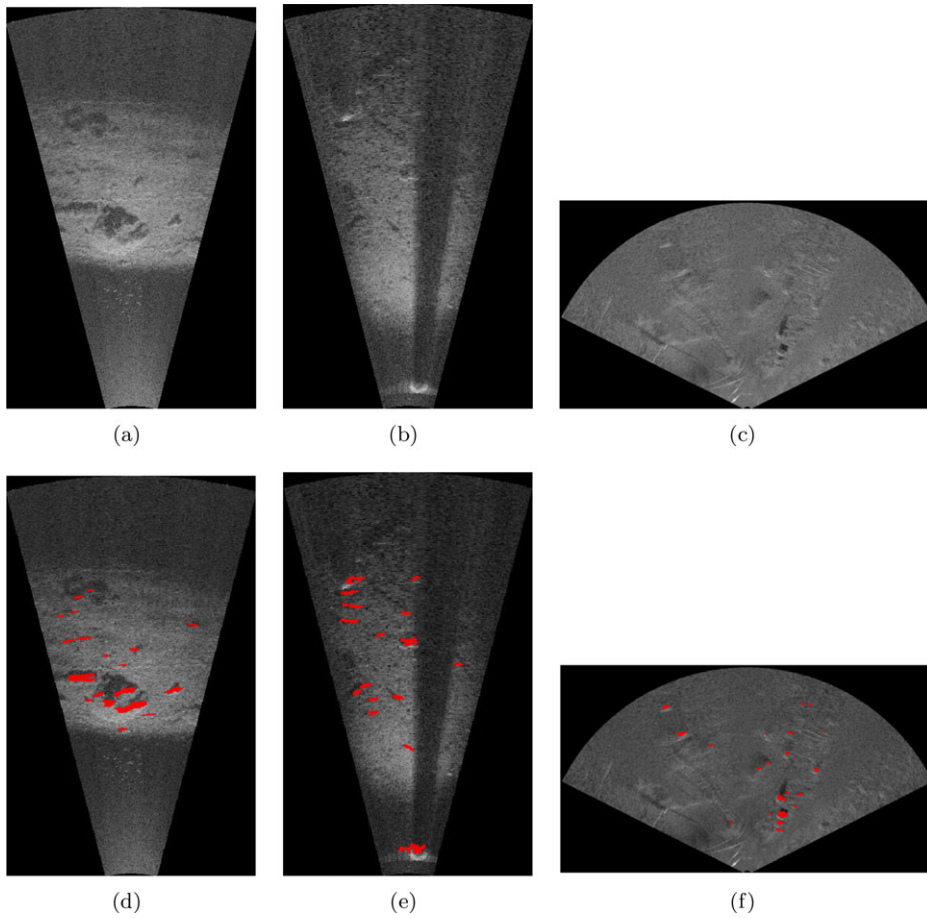
Tables III and IV summarize the results of each test, showing the mean and maximum errors for the rotation and translations in each dataset.

We start by analyzing the registration of consecutive frames. In the case of the first two datasets, both methods present low errors, with a slightly better performance by the Fourier-based method. The high resolution of the sonar together with the prominent features in the environment allow for an accurate estimation of alignments along the two sequences. The third dataset presents higher errors due to the lower resolution of the acquisition. The Fourier-based registration outperforms the region-based technique both in translation and rotation estimation. In general, the type of features in this dataset—sparser and weaker—is likely to generate unstable regions. However, since the images are spatially close, the error remains reasonably low.

Regarding the second test, in which the registered frames are more distant, we observe that, as expected, the results tend to have a higher error rate. In particular, the errors for the region-based method have especially increased with respect to their counterparts in the first experiment. When the features are initially far apart and a good initial prior is not available, the NDT algorithm may converge to a local minima, giving rise to erroneous estimations. Besides, in the first two datasets, the smaller overlap and the narrow aperture of the sonar in the azimuth direction (14.4°) cause significant features to drop out of the field of view eventually, leading to an insufficient number of features to perform the NDT alignment in a reliable manner. On the other hand, the content of the overlapping area, although smaller than in the first test, is sufficient to find the correct correlation with the Fourier-based method, thus yielding a lower mean error.

The errors in the third dataset have increased under both registration methodologies compared to the previous test. The error of the rotation estimation, which is the motion most affected by intensity alterations, is especially high. Note that the feature extraction algorithm targets the transitions from protruding objects to the shadows or the background plane. With the change in the sonar’s vantage point, these transitions can vary substantially, and therefore the extracted features from both views exhibit different layouts and cannot be correctly aligned. In our proposed method, since the information incorporated in the registration process is not only limited to the object transitions, other areas in the image can contribute to the anchoring to the correct registration point. The lower error of the Fourier-based technique when compared to the method of Johannsson *et al.* testifies to its better performance in these situations.

As all the analyzed sequences presented feature-rich environments, a different example is introduced to highlight



**Figure 5.** (a)–(c) Example frames of the datasets used for comparing the Fourier-based and the region-based registrations. (d)–(f) Example of extracted features with the method of Johannsson *et al.*

**Table III.** First test: Mean and maximum error of the registration in translation ( $t_x, t_y$ ) and rotation ( $\theta$ ) for the comparison experiments between the Fourier-based and region-based registration methods when registering consecutive frames.

	Region-based						Fourier-based					
	Mean error			Max error			Mean error			Max error		
	$t_x$ (m)	$t_y$ (m)	$\theta$ (deg)	$t_x$ (m)	$t_y$ (m)	$\theta$ (deg)	$t_x$ (m)	$t_y$ (m)	$\theta$ (deg)	$t_x$ (m)	$t_y$ (m)	$\theta$ (deg)
<b>Dataset 1</b>	0.11	0.06	1.01	2.91	1.63	9.56	0.09	0.06	0.51	1.29	0.23	0.61
<b>Dataset 2</b>	0.02	0.01	0.50	0.25	0.14	2.46	0.02	0.02	0.03	0.22	0.13	0.42
<b>Dataset 3</b>	0.44	0.42	1.08	13.9	8.33	14.6	0.23	0.15	0.54	3.20	2.25	7.60

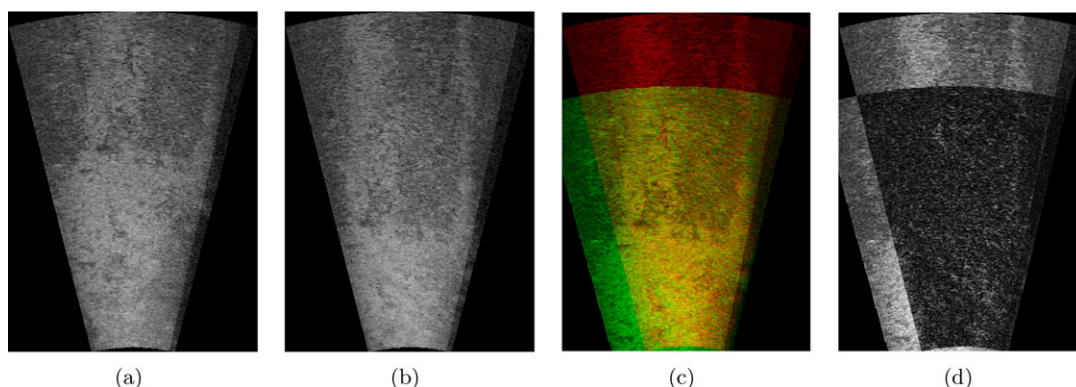
the difficulties of region-based techniques in environments with a scarcity of features. Figure 6 shows two images in a sequence lacking strong features. The method of Johannsson *et al.* is unable to extract any robust features as the thresholded negative gradients cannot be clustered in a sufficient number of points. On the other hand, the Fourier-based registration is able to align the views correctly by taking into

account the frequency information embedded in the different textures of the image. Although a ground-truth is not available, a composite overlay of two images in different color channels allows us to see that the correct alignment between the images has been found.

Hence, the proposed Fourier-based registration shows a superior performance in the alignment of both

**Table IV.** Second test: Mean and maximum error of the registration in translation ( $t_x, t_y$ ) and rotation ( $\theta$ ) for the comparison experiments between the Fourier-based and region-based registration methods when registering distant frames.

	Region-based						Fourier-based					
	Mean error			Max error			Mean error			Max error		
	$t_x$ (m)	$t_y$ (m)	$\theta$ (deg)	$t_x$ (m)	$t_y$ (m)	$\theta$ (deg)	$t_x$ (m)	$t_y$ (m)	$\theta$ (deg)	$t_x$ (m)	$t_y$ (m)	$\theta$ (deg)
<b>Dataset 1</b>	0.60	0.26	1.34	5.62	3.63	12.5	0.35	0.24	1.15	1.05	1.11	3.91
<b>Dataset 2</b>	0.45	0.21	1.03	1.63	4.93	18.5	0.11	0.23	0.09	0.22	0.13	5.52
<b>Dataset 3</b>	0.81	0.92	2.80	17.9	11.5	27.6	0.34	0.18	1.72	15.0	5.14	29.5



**Figure 6.** (a),(b) Example frames of a featureless dataset. (c),(d) Registration performed by the Fourier-based method: (c) Overlay of the two registered images in different color channels. Note the correct alignment in the yellow area. (d) Difference image of the registered frames. Note that almost all content in the registered area has been subtracted as a consequence of the alignment.

consecutive and nonconsecutive frames and higher robustness in difficult environments. As a result, the possibility of establishing registration constraints between two views is increased with the benefits that this implies.

The performance of the methods with regard to the computational complexity is also worth noting. The most demanding aspect of our proposed method is the computations of the Fourier transforms. The FFT algorithm requires  $O(2N^2 \log_2 N)$  operations for each 2D transform, where  $N^2$  is the number of pixels in the image.

The current implementation of the registration algorithm, coded in Python and making use of the ANFFT libraries (ANFFT Package, 2013), consumes approximately 60 ms per pairwise registration in an Intel i7 at 3.4 MHz, considering typical image sizes under  $1,024 \times 1,024$  pixels and a single-threaded execution. On the other hand, our implementation of the Johannsson *et al.* technique takes approximately six times longer, the major part of the time being consumed by the NDT optimization process.

#### 4. GLOBAL ALIGNMENT

The registration method described so far is intended to compute the relative transformation between pairs of overlapping images. To generate a mosaic, it is necessary to map all

the images into a common reference frame. This is normally accomplished by concatenating the transformations of successive images so that the transformation between nonconsecutive views is obtained. However, it is well known that chaining transformations over long sequences is prone to cumulative error (Smith & Cheeseman, 1986). With the aim of obtaining a globally consistent set of transformations, the problem is reshaped into a pose-based graph optimization. A least-squares minimization is formulated to estimate the maximum-likelihood configuration of the sonar images based on the pairwise constraints between consecutive and nonconsecutive registrations. As our main concern here is the mosaic generation, the problem is approached in an offline fashion. However, if the registration constraints were to be used in a motion estimation framework, they could be integrated into online back ends developed to efficiently optimize pose graphs, such as incremental smoothing and mapping (iSAM) (Kaess et al., 2008).

Two different situations are considered throughout this section: working exclusively with FLS imagery or also being able to incorporate navigation measurements from other sensor data. The high frame rate of FLS allows us to contemplate the case of dealing only with imagery, extending the applicability of the method to situations in which the sonar is deployed from vehicles with reduced sensor suites or

other situations in which acquiring correct navigation data might be difficult (e.g., using a compass close to magnetic disturbances).

We first present the general formulation of the pose-based graph followed by an explanation of how the uncertainty of the registration constraints is estimated. Finally, we describe the methodology to select which frame links are to be included in the graph in order to increase the efficiency of the global alignment step.

#### 4.1 Graph Definition

We define a graph whose vertices represent the position of observed sonar images and whose edges are pose constraints obtained from the pairwise registrations. Let  $\mathbf{v} = (\mathbf{v}_1, \dots, \mathbf{v}_n)^T$  be a set of vertices, where  $\mathbf{v}_i = (x_i, y_i, \theta_i)$  describes the position and orientation of sonar image  $i$ . When relying solely on image data, the initial positions of the vertices are estimated using the chained transformations between consecutive image pairs. If navigation data are available, the vertices can be initialized using the pose estimates from the dead-reckoning information.

Let  $\mathbf{z}_{i,j}$  and  $\Omega_{i,j}^z$  be the mean and information matrix, respectively, of the transformation from image  $i$  to image  $j$  obtained from applying the registration algorithm on the image pair  $(i, j)$ . Let  $\hat{\mathbf{z}}_{ij}(\mathbf{v}_i, \mathbf{v}_j)$  be the expected transformation given the configuration of  $\mathbf{v}_i$  and  $\mathbf{v}_j$ .

Then, we can define an error function of the following form:

$$\mathbf{e}(\mathbf{v}_i, \mathbf{v}_j, \mathbf{z}_{i,j}) = \mathbf{z}_{i,j} \ominus \hat{\mathbf{z}}_{ij}(\mathbf{v}_i, \mathbf{v}_j), \quad (8)$$

where  $\ominus$  is the inverse of the usual motion composition operator in the 2D Euclidean space.

Essentially, the error function measures how well the position blocks  $\mathbf{v}_i$  and  $\mathbf{v}_j$  satisfy the constraint  $\mathbf{z}_{i,j}$ . Therefore, to find the most consistent spatial arrangement for all the image poses, we seek the configuration of the vertices  $\mathbf{v}^*$  that minimizes the negative log likelihood of the set of all existing constraints  $\mathcal{C}$ :

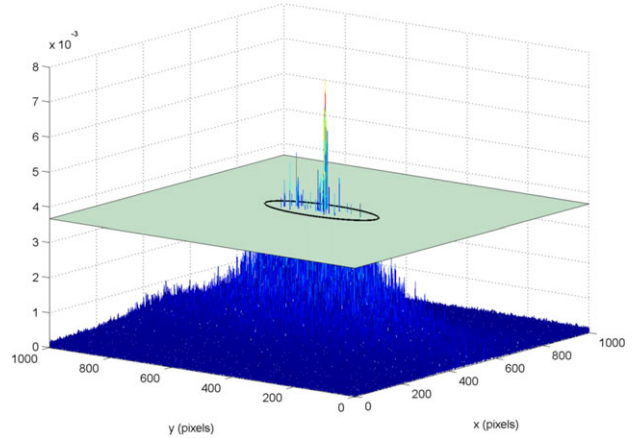
$$F(\mathbf{v}) = \sum_{(i,j) \in \mathcal{C}} \mathbf{e}(\mathbf{v}_i, \mathbf{v}_j, \mathbf{z}_{i,j})^T \Omega_{i,j}^z \mathbf{e}(\mathbf{v}_i, \mathbf{v}_j, \mathbf{z}_{i,j}), \quad (9)$$

$$\mathbf{v}^* = \arg \min_{\mathbf{v}} [F(\mathbf{v})]. \quad (10)$$

If a good initial guess of the parameters is known, a numerical solution of Eq. (10) can be obtained using the popular Levenberg-Marquardt algorithm (Moré, 1978). In our implementation, this minimization is solved using the General Framework for Graph Optimization (g2o) back end (Kummerle et al., 2011).

#### 4.2. Estimation of the Registration Uncertainty

The described pose-graph formulation requires establishing an information matrix  $\Omega^z$  for every registration measurement. To this end, a heuristic is derived from the registra-



**Figure 7.** Representation of the proposed heuristic to compute the uncertainty of the registration from the phase correlation matrix values. Ellipse represented using a confidence interval of 99%.

tion method in order to quantify the degree of confidence in the alignment. Recalling the description of the method in Section 3.1, the values of the phase correlation matrix can be used as a direct measure of the degree of congruence between two images. The amplitude and extent of values surrounding the main peak account for localization inaccuracies in the registration.

Therefore, the phase correlation surface is thresholded at a given amplitude, and the standard deviations of the  $x$  and  $y$  coordinates of the matrix cells that exceed the threshold are extracted, as depicted in Figure 7. The threshold is set to half the power of the main peak.

This procedure is applied to the phase correlation matrices obtained from both the rotation and the translation estimation steps, resulting in three different uncertainties  $(\sigma_x, \sigma_y, \sigma_r)$ . These values, obtained in pixels, are then converted to meters and radians by using the range resolution of the sonar  $\delta_r$  (pixels/m) according to the experiment's configuration and the angular resolution of the polar sonar images  $\delta_\theta$  (pixels/rad). Finally, the values are reshaped in a covariance matrix, which is inverted, yielding the information matrix  $\Omega^z$  of the measurement:

$$\Omega^z = \begin{bmatrix} (\sigma_x \delta_r)^2 & 0 & 0 \\ 0 & (\sigma_y \delta_r)^2 & 0 \\ 0 & 0 & (\sigma_r \delta_\theta)^2 \end{bmatrix}^{-1}. \quad (11)$$

A similar heuristic was proposed in earlier work by Pflingstorn, Birk, Schwertfeger, Bülow, and Pathak (2010). Their heuristic fits a  $2 \times 2$  covariance matrix to a window of size  $K$  around the registration result (i.e., the main peak of the correlation matrix). The heuristic weights the squared distance to the mean of the values inside the window by the normalized amplitudes of the phase correlation. The

**Table V.** Evaluation of the heuristics to estimate the uncertainty of the registration: average distance errors with respect to the GPS ground truth of absolute trajectories obtained with each method.

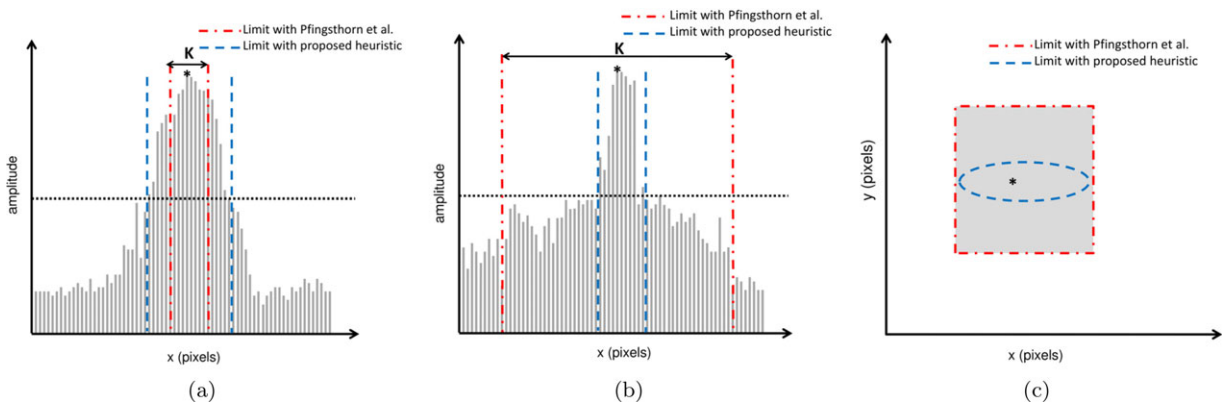
Heuristic	Average distance error (m)
Proposed	10.2
Pfingsthorn et al., $K = 250$	27.8
Pfingsthorn et al., $K = 500$	13.4
Pfingsthorn et al., $K = 1,000$	12.9

outcome is then strongly dependent on window size  $K$ , although how this value is selected is not shown in their experiments, nor are the typical values for this parameter reported. Contrary to this, our strategy readily offers a way to adapt the values that contribute to the variance computation by taking into account those values that are above half power of the main peak (i.e., values within 3 dB below the peak).

Since we do not have the means of computing the true uncertainty of a registration, it is difficult to assess the performance of the proposed heuristic against that of Pfingsthorn et al. The solution we adopt is to use a dataset with an available ground truth to evaluate the final trajectory result obtained by using each of the heuristics. To this aim, we have selected a portion of the Marciana Marina dataset presented in Section 6.3, in particular from frames 250 to 1,000. We have computed the registrations among all these frames and we have built two graphs, one using

each heuristic in the computation of the uncertainties. After the optimization, the estimated trajectory is compared with the ground truth path. The error between these two is indicative of which heuristic leads to a better description of the graph uncertainties, thus leading to a solution that converges more closely to the real one. The uncertainties of the method of Pfingsthorn et al. have been computed for three different values of  $K$ . The values have been selected by taking into account the dimensions of the phase correlation matrix, which for the dataset’s images is  $1,526 \times 1,526$ . In this way, we have chosen small  $K = 250$ , medium  $K = 500$ , and large  $K = 1,000$  values. Table V summarizes the absolute mean distances for each of the trajectories with respect to the ground truth, computed by averaging the distance of all nodes to their corresponding ones in the ground truth trajectory.

A logical explanation for these results might be found by considering the implication of the window size parameter. If  $K$  is too small, the obtained uncertainty measures might eventually be limited to values that do not represent the uncertainty of the main peak, leading to too optimistic uncertainty measures [as depicted in the schematic in Figure 8(a)]. That would explain the high values obtained with the small  $K$ . On the other hand, large  $K$  values are a better strategy, given that covariances are weighted by the corresponding intensities (which are expected to be low if the values are far apart from the main peak). In this case, even if a high number of values take part in the computation, they would have a low weight in it. However, if the values located far from the main peak do not have such low intensities (as may happen in noisy sonar images where the correlation matrix has a lot of scattered noise peaks), it could



**Figure 8.** Schematics illustrating the performance of the proposed heuristic for estimating the registration uncertainty versus the Pfingsthorn *et al.* heuristic. (a) Example of a small  $K$  value that is not able to represent the uncertainty of the registration peak. (b) Example of a situation in which Pfingsthorn *et al.*'s heuristic would provide a pessimistic uncertainty as a consequence of all the small contributions inside the  $K$  window. (c) Scheme of a phase correlation matrix seen from the top view, illustrating a typical case in which the main peak is spread in one direction as a consequence of the motion direction. The drawn limits represent the values that would be considered for the variance computation: while our method would consider only the ones over the threshold, thus adapting to the approximate elliptical shape of the main peak, the method of Pfingsthorn et al. would use the squared window, and therefore all the values in the gray area would also contribute to increasing the uncertainty in the  $y$  direction.

lead to an overpessimistic computation of the uncertainty in some cases [as illustrated in the schematic in Figure 8(b)]. Likewise, the fact of considering a squared window may also lead to overpessimistic estimates [Figure 8(c)]. If the shape of the main peak has, for instance, an elliptical contour (as is common under one-directional displacements where the peak is smeared in the motion direction), a large number of contributions will unnecessarily increase the uncertainty in the other direction (even though their weight in the computation is small). This might be the explanation for the larger errors for  $K = 500$  and 1,000. On the other hand, our proposed technique takes into account only the peaks surpassing the half power threshold and that are significant enough to influence the registration result, independently of how they are spatially arranged.

It is worth highlighting that the heuristic of Pflingstorn et al. was conceived to estimate the uncertainty of phase correlation registrations over optical images, which usually suffer from less noise and fewer artifacts than their sonar counterparts. In these cases, correlation peaks are narrower and the heuristic is not affected by the aforementioned issues, thus definitely being a good strategy to estimate the uncertainty. However, in the case of FLS images, correlation matrices present smeared main peaks and more scattered noise. For this reason, we chose to apply the proposed heuristic to measure the registration uncertainty.

### 4.3. Link Candidates

To avoid unnecessary computations, it is essential to attempt registration only with frame pairs that are likely to overlap. To detect these candidate pairs, and particularly in the case of nonconsecutive overlapping images, it is necessary to first infer the path topology.

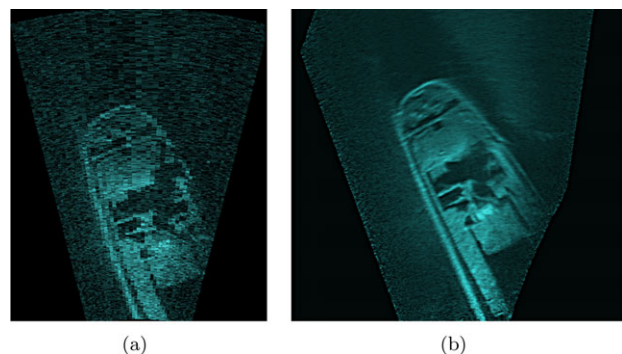
In the absence of other sensor data, the path topology is inferred by using the registrations of consecutive images. We compute the registrations of each frame with their successive but also with several of their neighboring frames by establishing a fixed window around the current sequence position. The size of this window is estimated according to the range and mean velocity of the sonar so as to select sequential frames going from the next neighbor down to a frame with approximately 50% overlap. The computation of these multiple links helps to increase the local robustness of the initial estimated path. The obtained constraints, together with their uncertainties, are fed to the graph optimization back end and an initial estimation of the path is obtained. Notice that only the registrations with a low uncertainty (according to an established threshold) will be introduced as constraints in the graph. Otherwise, if navigation data are available, the initial path comes readily from the dead-reckoning estimates. With this initial guess, putative overlapping pairs can be identified according to the spatial arrangement of the image's positions. To check the overlap between two images, their fan-shaped foot-

prints are projected over a plane according to their position and orientation. The ratio of the intersection area over the total area of the footprint is computed, and two criteria are imposed for it to be considered a valid candidate pair: first, the overlap percentage must be above an established threshold, and second, the orientation difference between the two frames must fall within the limits  $[-\frac{\text{FoV}}{2} : \frac{\text{FoV}}{2}]$  or  $[-\frac{\text{FoV}}{2} + 180 : 180 + \frac{\text{FoV}}{2}]$ . In this way, we avoid selecting as candidate pairs those frames that, even presenting enough overlap, cannot be registered given the implicit restrictions of our rotation estimation algorithm. Note that in an on-line approach, in order to maintain a consistent estimation throughout the time, the overlap checking would have to be performed at each new frame (or at each group of  $n$  new frames) in order to identify possible loop closures with the frames previously incorporated into the graph.

Once the candidate pairs are identified, they are fed into the registration module described in Section 3.1. Finally, in order to avoid introducing erroneous constraints in the graph, a second level of pruning is performed to discard, from all the attempted registrations, those with large uncertainty. This filtering is performed according to an established threshold on the uncertainty measure described in the previous section for both rotation and translation estimates.

## 5. MOSAIC RENDERING

The global alignment provides the position of the sonar images in a global reference frame, usually the first frame of the image sequence. We can then build the absolute homographies that relate every image with respect to the reference frame and map the sonar images on the mosaic plane. However, as the content of multiple images will overlap in a given position, a strategy is required to deal with the



**Figure 9.** Example of the denoising effect obtained by mosaicing. (a) Single frame gathered with a DIDSON sonar operating at its lower frequency (1.1.MHz). (b) Small mosaic composed of 50 registered frames from the same sequence blended by averaging the overlapping intensities. It can be clearly seen how the SNR increases and the details pop out.



combination of the pixel intensities and thus generate a representation with a smooth and continuous overall appearance.

Again, we cannot take advantage of traditional blending techniques designed for video images. Optical blending generally deals with a low number of images at a given position by treating only the intersecting boundaries. Some methods focus on finding the optimal location to place a seam that minimizes the photometrical and geometrical changes, whereas others apply a smooth transition over the intersection area. However, blending an FLS mosaic requires dealing with multiple images due to the high overlap percentages. Especially when the images have been acquired in an across-range fashion, high overlap is a must to achieve good coverage due to the sonar fan-shaped footprint. Furthermore, presuming that a correct registration has been performed, it is of interest to keep as much of the overlapping images as possible to be able to improve the SNR of the final mosaic. Therefore, it is necessary to deal not only with the seam areas, but with the whole image.

A simple but effective strategy is to perform an average of the intensities that are mapped to the same location. Averaging the overlapping sonar intensities yields the denoising of the final mosaic, achieving an improvement in terms of SNR compared to a single image frame (see Figure 9). Ideally, by averaging, the reduction of the noise (and therefore the improvement of the SNR) is proportional to the squared number of averaged samples. Then, under the assumption of additive Gaussian noise, a mosaic would have an overall SNR improvement of approximately the mean of the square roots of the overlapped images at each location. However, we must highlight the fact that averaging reduces only the contributions of random uncorrelated noise, and therefore the image SNR cannot be increased indefinitely by averaging more samples as, eventually, the remaining noise is due to artifacts that may manifest as correlated noise. In the presence of registration misalignments, the averaging strategy will generate blurred areas of mixed content. Besides, several photometrical artifacts may arise, such as noticeable seams due to a nonconstant number of overlapping images, especially in datasets with different tracklines, rotations, or a nonconstant vehicle speed. A fade-out of the mosaic content can also occur when averaging images that have blind (i.e., black) areas, typically as a consequence of nonproper imaging configurations. In Hurtós et al. (2013a), we have proposed a blending workflow that preserves the averaging nature but enables the correction of these artifacts through different strategies. Although more sophisticated blending techniques are being investigated, we choose to use here the average blending as it also serves as a visual indicator of the mosaic's consistency.

Apart from the improvements in SNR, the resolution of the final mosaic can also be enhanced with respect to the original images. We can take advantage of the multi-

ple alignment of low-resolution images together with the subpixel accuracy positions obtained from the global alignment step to perform superresolution. Hence, by oversampling the mosaic grid and mapping the images with subpixel transformations, we achieve a higher resolution and an overall enhancement of the mosaic image.

## 6. EXPERIMENTS AND RESULTS

This section presents experimental results, validating both the proposed registration methodology and the global alignment strategy. We first report on a small test performed inside a water tank, followed by results on relevant field applications that take place inside harbor environments where visibility is often compromised.

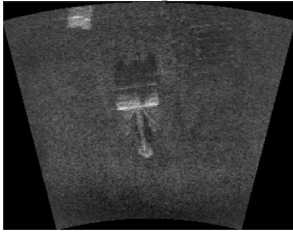
For each experiment, we report the details of the input sonar frames, the computation times, and the resolution and size of the obtained mosaics (Tables VI, VII, and VIII). All the computation times are obtained using an Intel i7 3.4 MHz QuadCore CPU. Notice that, even when we run the experiments in a CPU with multiple cores, the implementation of our registration algorithm, as stated in Section 3.2, is running in a single core. A more efficient multithreaded implementation is under development.

### 6.1. Tank Test

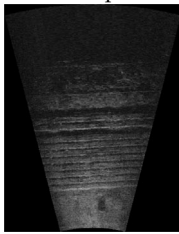
This experiment was carried out in the water tank at the University of Girona using the Girona-500 AUV (Ribas, Palomeras, Ridao, Carreras, & Mallios, 2012), equipped with the ARIS FLS. Several objects were deployed at the bottom of the tank and the vehicle was teleoperated over them while maintaining the same orientation viewpoint throughout the experiment. Due to the limited size of the tank, the sonar was configured for small ranges, imaging a window of 1.5 m. A total of 527 frames were acquired, with the vehicle navigating at a constant altitude of 1.5 m from the bottom and the sonar tilted at  $20^\circ$  to facilitate good imaging conditions. Ground truth is not available as the indoor environment of the experiment did not allow the use of a GPS.

Figure 10(a) shows the vehicle's dead-reckoning trajectory together with several of the estimated trajectories. The black dashed line shows the estimated trajectory computed by concatenation of the registration constraints of consecutive image pairs. The green dashed line shows the estimated path, including constraints computed within a window of 20 neighboring frames. It can be seen that with only the incorporation of these local constraints, the solution comes much closer to the final global-aligned trajectory (depicted in red). This green path has been used as the initial guess for the strategy to find a link hypothesis. Due to the high overlap of the sequence, the method returned 36,092 potentially overlapping pairs under the requirement of a 50% overlap. From these, 15,473 were considered successful registrations according to strict uncertainty thresholds.

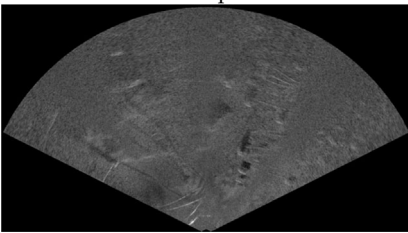
**Table VI.** Summary table for the Tank Test Dataset.

TANK TEST	Input frame	Size (pixels)	Resolution (m/pixel)	Example
		350 × 274	0.0045	
	Computation times	Registration(s) 2312	Optimization(s) 27.6	
	Final mosaic	Size (m) 2.3 × 3	Resolution (m/pixel) 0.0011	

**Table VII.** Summary table for the ship hull dataset.

SHIP HULL	Input frame	Size (pixels)	Resolution (m/pixel)	Example
		350 × 453	0.01	
	Computation times	Registration(s) 2654	Optimization(s) 31.3	
	Final mosaic	Size (m) 19 × 9.5	Resolution (m/pixel) 0.003	

**Table VIII.** Summary table for the Marina Marciana dataset.

MARCIANA MARINA	Input Frame	Size (pixels)	Resolution (m/pixel)	Example
		1,526 × 848	0.057	
	Computation times	Registration (s) 10316	Optimization (s) 219	
	Final mosaic	Size (m) 512 × 352	Resolution (m/pixel) 0.057	

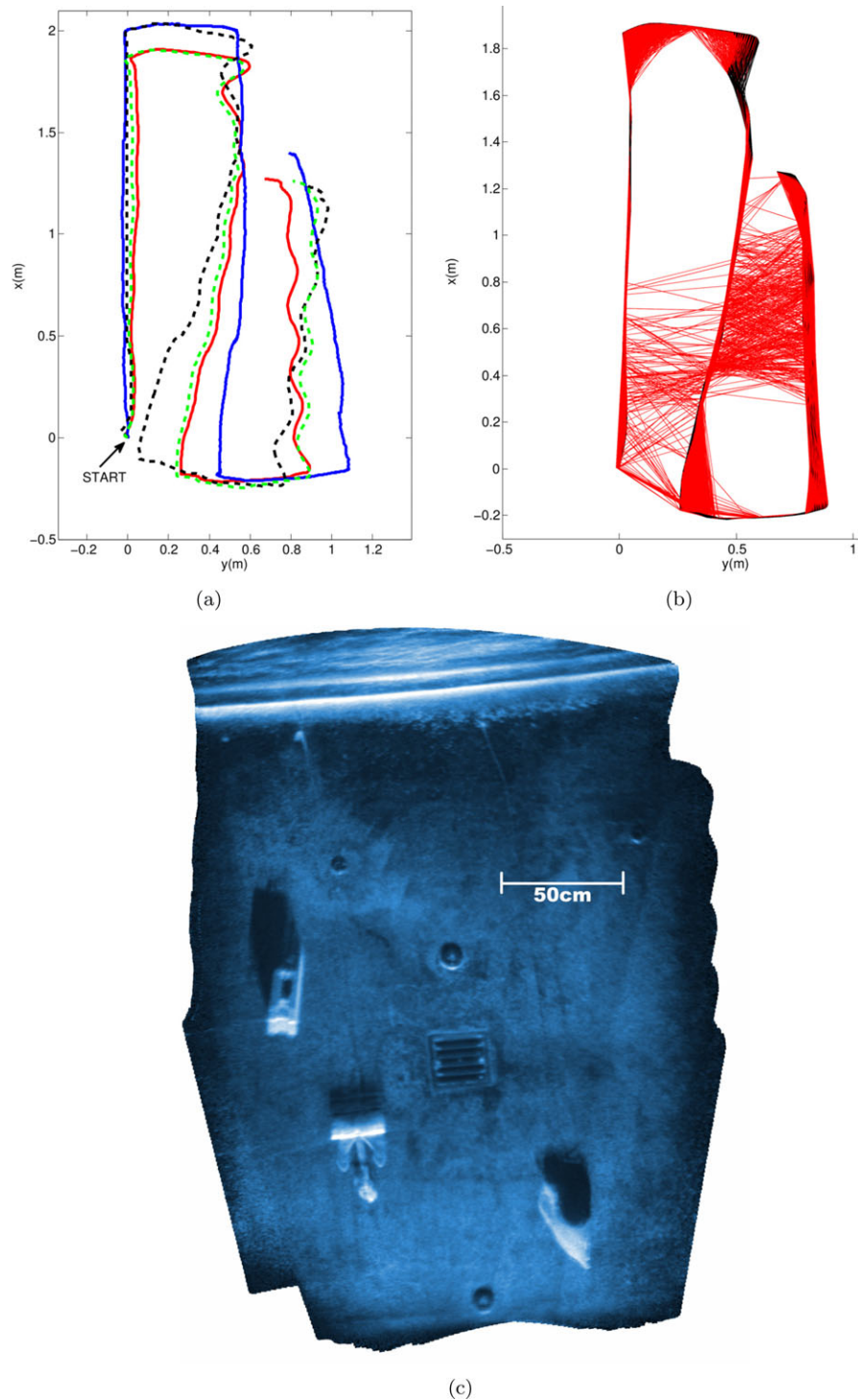
Given the high resolution delivered by the ARIS sonar, i.e., an angular resolution of  $0.2^\circ$  and a range resolution of 5 mm/pixel in the experiment configuration, the uncertainty thresholds for translation and rotation were established at 2 cm and  $2.5^\circ$ , respectively. Figure 10(b) shows the final constraints included in the graph. Due to the high overlap of the dataset images, the graph presents a large number of constraints. While this does not present any difficulties in an offline framework, better pruning of the candidates would be necessary to solve the problem in real time.

Figure 10(c) shows the obtained mosaic composed of 527 frames and rendered over an oversampled grid at four times the original resolution. The mosaic shows high self-consistence, supporting the accuracy of the method, and it enables the identification of the small objects present in the scene: a concrete block, an anchor, and an amphora, as well as a grill and other details of the tank. Figure 11 shows a

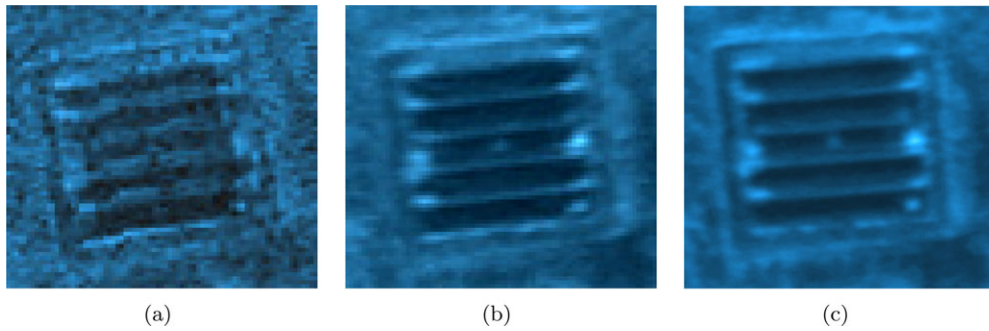
comparison of a detailed area where the SNR improvement between a single sonar frame [Figure 11(a)] and the mosaic [Figure 11(b)] can be easily appreciated. Note also the enhancement of the image when comparing the oversampled [Figure 11(c)] and nonoversampled [Figure 11(b)] versions of the mosaic.

## 6.2. Ship Hull Inspection

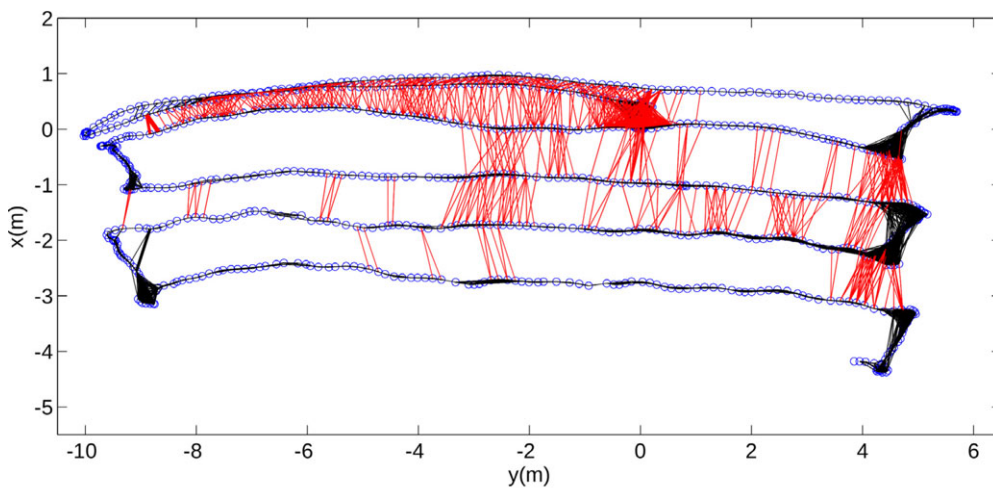
The second experiment is in the context of ship hull inspection. The dataset was acquired with the Hovering Autonomous Underwater Vehicle (HAUV) (Vaganay *et al.*, 2005) of Bluefin Robotics (Bluefin Robotics Corp., 2013) and a DIDSON sonar. The vehicle navigated across the bottom of a ship hull, maintaining a constant distance to it and covering an area of about  $15 \text{ m} \times 6 \text{ m}$ . The sonar was mounted on a tilt unit and was actuated throughout the experiment



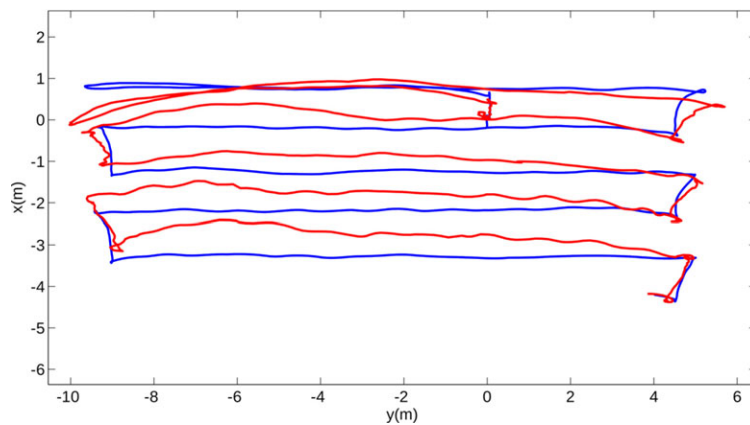
**Figure 10.** (a) Trajectories of the ARIS Tank experiment. Blue: Vehicle’s dead-reckoning trajectory. Black-dashed: Trajectory estimation from consecutive image registrations. Green-dashed: Trajectory estimation from the consecutive constraints including a window of local neighbors. Red: Final estimated trajectory after the global alignment. (b) Final graph constraints of the ARIS Tank experiment. Black: window constraints. Red: loop-closure constraints. (c) Mosaic composition with the 527 frames from the ARIS Tank experiment.



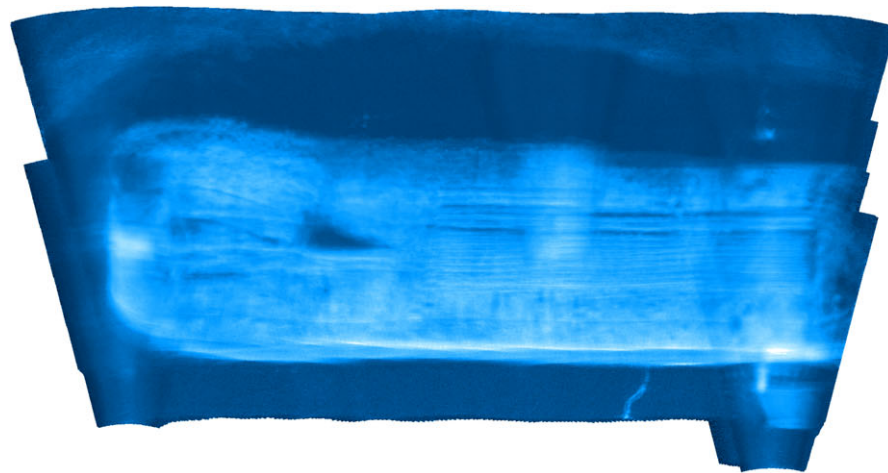
**Figure 11.** Detailed comparison of a small area in the tank mosaic. (a) Single frame. (b) Nonoversampled mosaic. (c) Mosaic oversampled four times the original resolution. Note the improvement of the mosaic SNR with respect to the individual frame and the enhancement of the oversampled mosaic version.



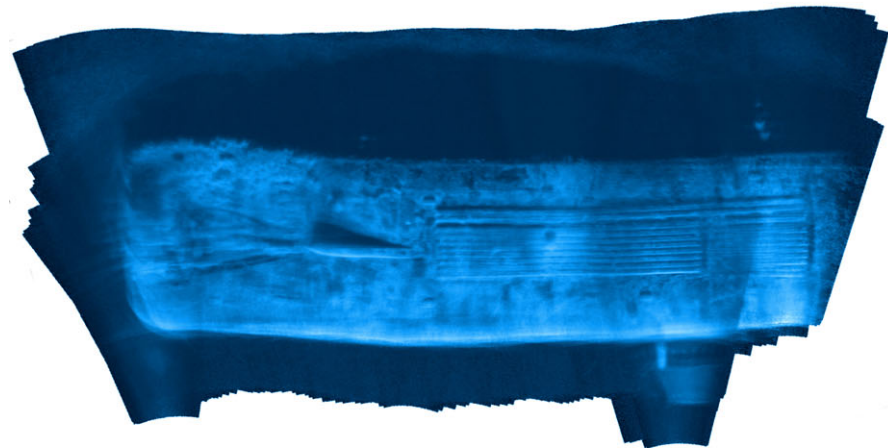
**Figure 12.** Links established by registration constraints in the ship hull dataset. Blue circles represent the vertices of the graph. Links in black depict the registration of a frame with neighboring frames inside a window. Links in red represent constraints found in loop-closure situations.



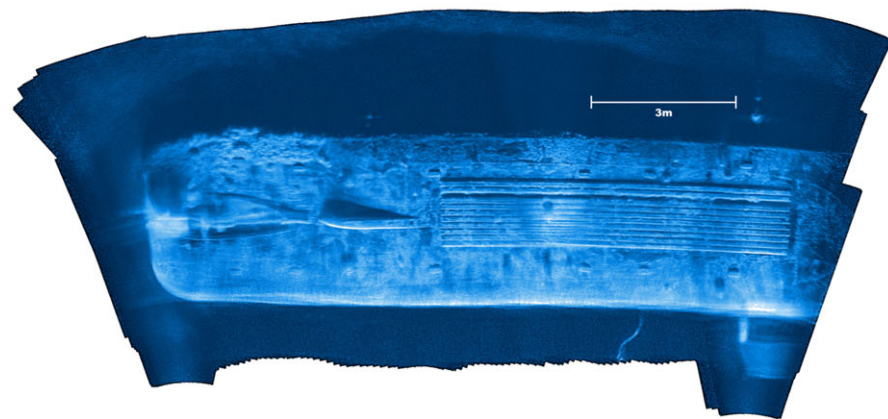
**Figure 13.** Trajectories of the ship hull dataset: Navigation trajectory (in blue) and estimated trajectory after the global alignment (in red).



(a)

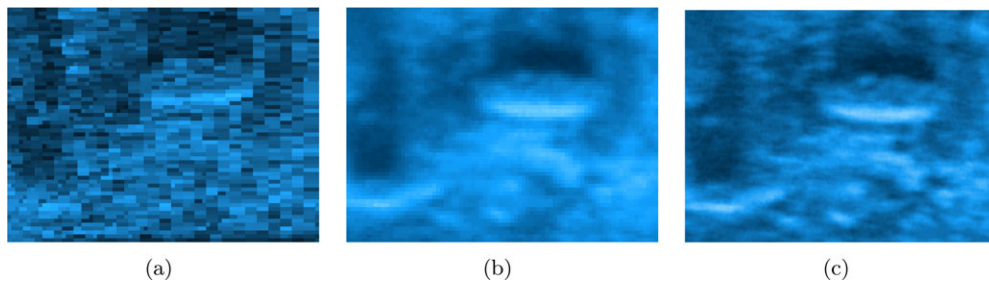


(b)



(c)

**Figure 14.** Ship hull mosaic rendered over different trajectories: (a) Over navigation trajectory. (b) Over the estimated trajectory before the optimization. (c) Over the final optimized trajectory. Parts (a) and (b) present blurred areas as a consequence of averaging misaligned images, whereas the final mosaic shows high consistency.



**Figure 15.** Detail comparison of a small area in the ship hull mosaic. (a) Single frame. (b) Nonoversampled mosaic. (c) Mosaic oversampled at three times the original resolution. Note the improvement in the mosaic SNR with respect to the individual frame and the enhancement of the oversampled mosaic version.

so as to adapt the images to the hull's surface and facilitate better imaging conditions. The final trajectory consists of five tracklines across the bottom of the hull, comprising a total of 4,420 sonar images collected during 7 min. The spacing between the tracklines (about 1 m) and the range configuration of the sonar (up to 4.5 m) guarantees sufficient overlap between different tracks. Moreover, the vehicle was moving basically in surge and heave degrees of freedom, which facilitates the registration between revisited locations as the vantage point is preserved throughout the experiment.

Due to the high frame rate of acquisition (six frames per second), only one out of three images has been considered, reducing the dataset to 1,473 frames. No navigation information has been used in the global alignment stage. Following the link candidate strategy, a total of 17,079 pairwise registrations have been attempted, including frames from the vicinity of the sequence and frames found in loop closure situations. From these, 8,148 have been deemed correct registrations and therefore involved in the generated pose-based graph. The high number of established constraints (consecutive and nonconsecutive) allows us to obtain a consistent solution relying solely on the information extracted from the registrations. The total of performed registrations were computed in 22 min, which suggests that with a more restrictive pruning on the attempted frames, the algorithm can achieve real-time capabilities.

Figure 12 shows the final computed locations of the sonar images, depicting all the links established between frames. Figure 13 shows the navigation trajectory (in blue) referenced at the sonar's origin and the trajectory computed with our methodology (in red). Unfortunately, as the ground truth is not available in this dataset, we cannot provide a quantitative measure of which trajectory is better. However, by mapping the sonar frames over the image locations in both trajectories, one can appreciate that the mosaic over the estimated trajectory leads to a much more defined image composition. Figure 14(a) shows the mosaic over the navigation trajectory, while Figure 14(b) displays the mosaic of the estimated trajectory prior to the optimization step, and Figure 14(c) shows the final obtained mosaic oversam-

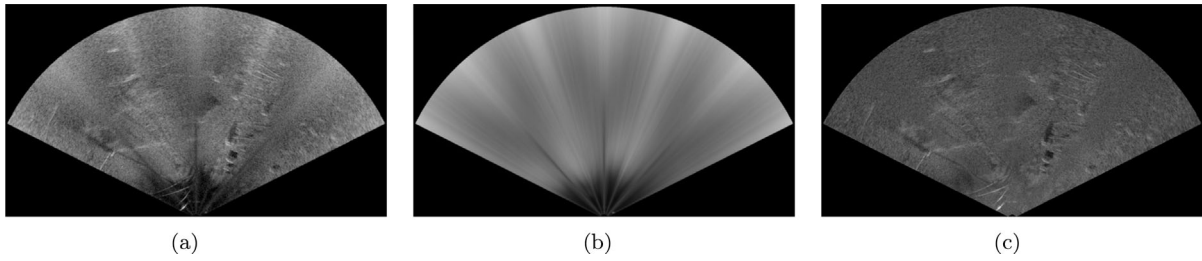
pled by a factor of 3. It can be seen that the composite image in Figure 14(c) presents a consistent overall appearance and allows the identification of the various features on the hull's bottom. Some illumination artifacts are present (especially in the lower part of the image) due to the tilt imaging angle.

Figure 15 shows a comparison of a detailed area between a single sonar frame, the nonoversampled version of the mosaic, and the mosaic oversampled at three times the original resolution.

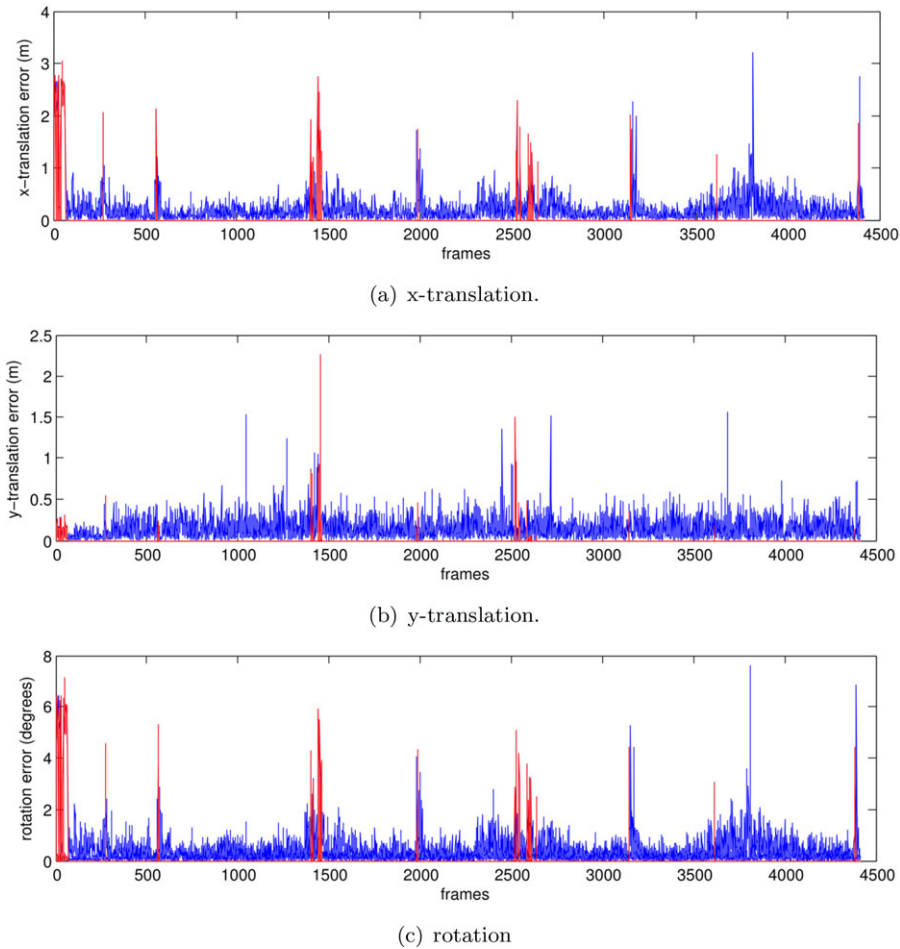
### 6.3. Marciana Marina Harbor

The third experiment is based on a harbor survey performed during the ANT'11 sea trial organized by the Centre for Maritime Research and Experimentation (CMRE), the former NATO Undersea Research Centre, located in La Spezia (Italy), during which the University of Girona collaborated with CMRE. The dataset was obtained using a Blueview P900-130 FLS mounted on CMRE's Catamaran Autonomous Surface Vehicle (a modified vessel made by Sea Robotics). The employed setup allows us to have precise differential GPS data and heading from 2 GPS units, which is used as ground truth. The dataset is composed of 4,416 sonar frames gathered along a 2.1 km trajectory comprising both translational and rotational motions. This dataset is useful to test the proposed methodology under some different conditions. The data are gathered in a natural environment containing typical seafloor features (e.g., vegetation, rocks) which are sparse and less prominent than those found in manmade scenarios. The acquisition sonar also has significant differences in its operating range (up to 50 m), field-of-view ( $130^\circ$ ), and resolution (5.8 cm/pixel) compared to the other reported experiments. Additionally, the frames present a strong inhomogeneous insonification pattern that has been corrected in a preprocessing step (Figure 16).

Although information from the GPS positions is available, it is not utilized to initialize the vertices of the pose graph. Given its high accuracy, it would result in an initial guess too close to the final solution and would prevent



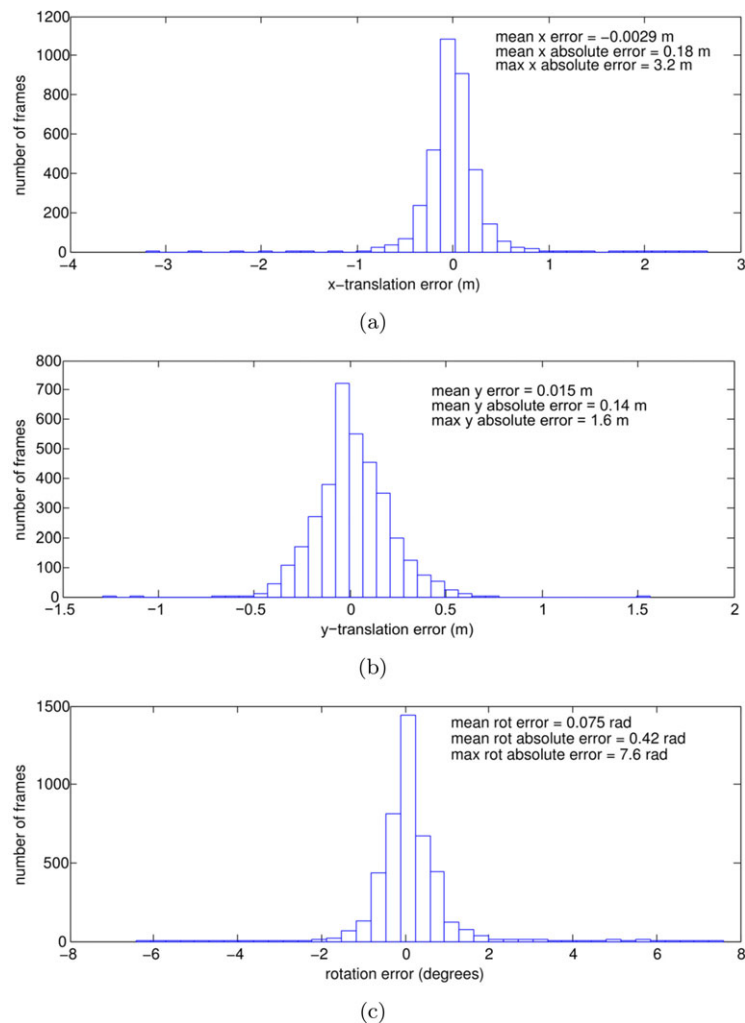
**Figure 16.** Image correction steps: (a) Frame example affected by inhomogeneous insonification. (b) Estimated illumination pattern. (c) Corrected frame.



**Figure 17.** Absolute mean errors (in  $x$  and  $y$  orientation) of the registration estimates for consecutive frames in the Marciana Marina dataset. Overlaid in red: errors of registrations that have been deemed unsuccessful according to the established thresholds on the uncertainty measure.

us from demonstrating the performance of the constraints established by the registration method. The proposed registration algorithm is generally successful in aligning the sequential image pairs of the dataset. Figure 17 shows the absolute mean errors of the registration estimates for con-

secutive frames compared to the ground truth odometry computed from the GPS positions. The mean errors are low, being 0.23 and 0.15 m for the  $x$  and  $y$  translations and 0.5° degrees for the rotational estimates. Colored in red, we depict those consecutive registrations that have been identified as



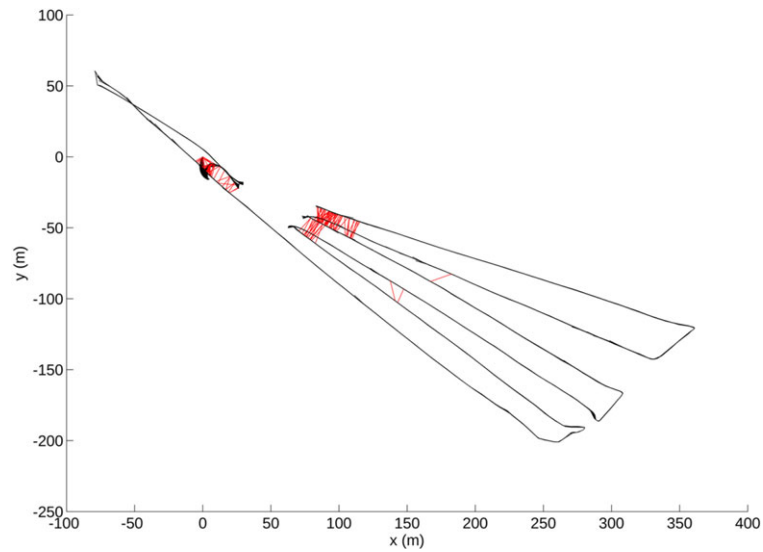
**Figure 18.** Error histograms for consecutive frame registrations in the Marciana Marina dataset. Only registrations considered successful under the established uncertainty threshold are taken into account. The maximum absolute errors are small and the mean of the error is around 0, indicating that the estimations are not affected by any bias. (a) Error histogram for  $x$ -translation. (b) Error histogram for  $y$ -translation. (c) Error histogram for orientation.

unsuccessful according to the thresholds on the uncertainty measure. Most of the frames with high error have been identified, and therefore are not introduced in the graph. High errors are arising mainly at the start of the sequence and around frames 1,500 and 2,500. Leaving aside the initial differences, where we believe the GPS had an issue with the heading, the last two problematic points correspond to two turns on the homogeneous areas, as can be seen in the lower right side of the mosaiced area [see Figure 21(a)]. In these areas, the images are highly homogeneous, lacking any type of content or intensity variation that causes the registration method to fail (for the rotation estimation, and subsequently for the translation).

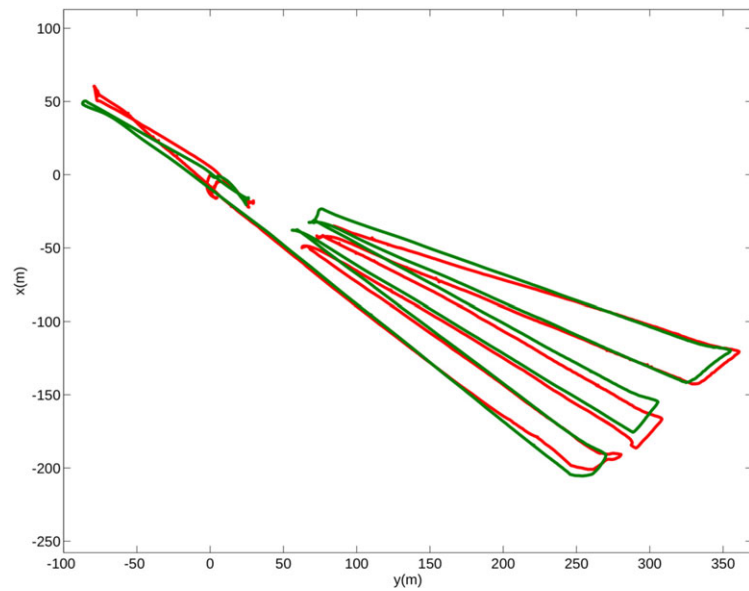
Figure 18 shows the error histograms of the registrations that have been identified as successful. As can be seen, the mean errors are around 0, indicating that the registration method is not affected by any bias. This is significant, as a bias in the registration estimates would not be addressed by the proposed optimization scheme.

The inability to link all consecutive frames prevents the generation of an initial graph using only the image information. In a sonar navigation framework, the dead-reckoning estimates would allow constraints to be established between these sonar poses. Here, we introduce in its place constraints based on the GPS measurements. Note that these constraints are only introduced in 64 of the 4,415 pairs of





**Figure 19.** Loop closure links detected in the Marciana Marina experiment.

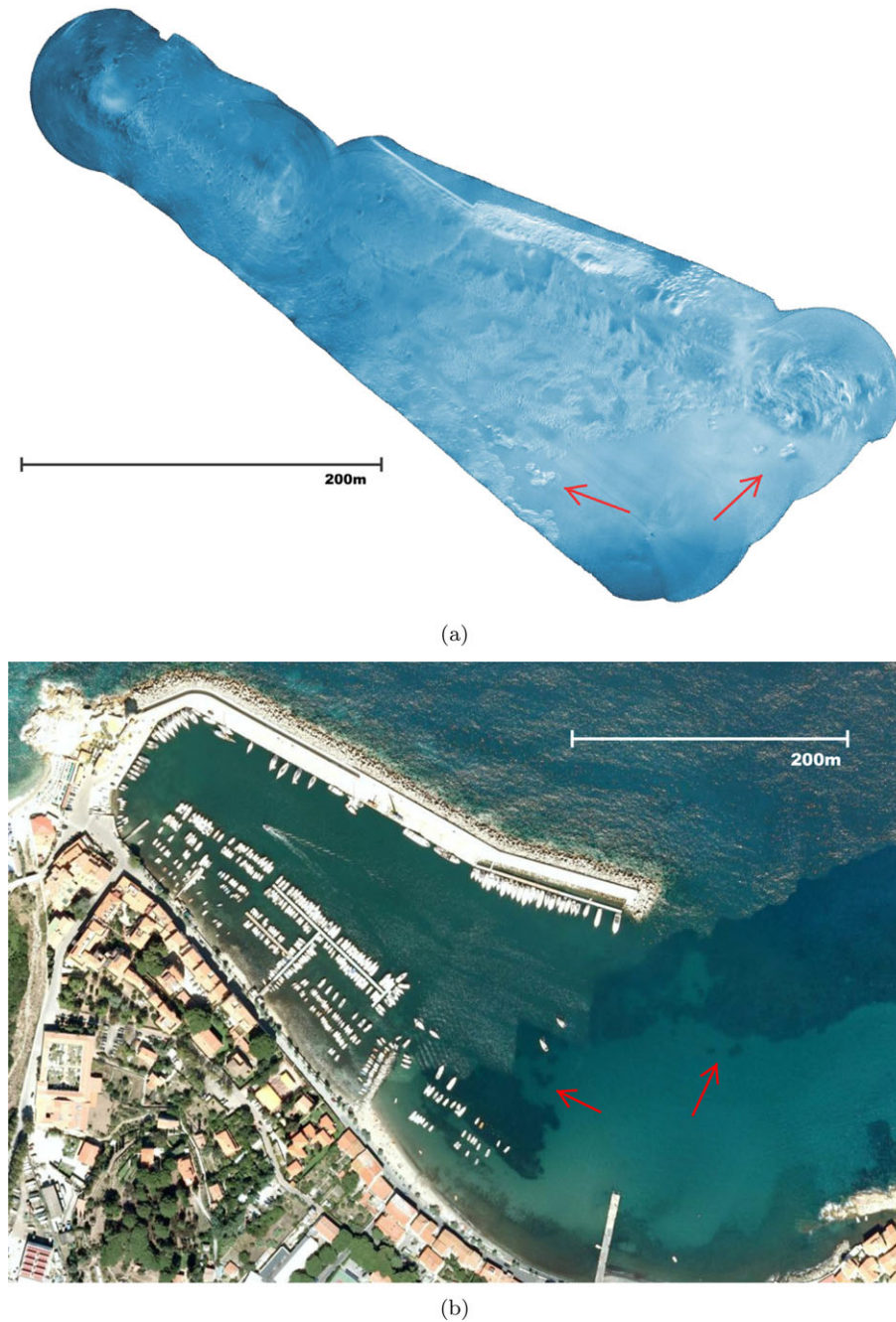


**Figure 20.** Trajectories of the Marciana Marina experiment: GPS trajectory (in green) and estimated trajectory after the global alignment (in red).

consecutive frames. These links, together with local links from the registrations inside a window of 20 neighboring frames, are used as an initial guess to determine a hypothesis for overlapping frames. In this experiment, cross-track registration is difficult since the vehicle navigated on nearly reciprocal headings, alternating them in consecutive tracks. That causes the image's appearance to suffer from significant changes and drastically lowers the number of detected loop closures, yet the registration algorithm is able to align

a small number of revisited frames crucial to enforce global consistency (Figure 19).

Figure 20 shows the GPS trajectory (in green) together with the estimated sonar trajectory (in red). It can be seen that the trajectory obtained after the graph optimization closely matches the GPS track, indicating that the registration constraints lead to a valid solution. There is a difference as a consequence of several small registration errors accumulated along the estimated path. These errors are



**Figure 21.** (a) Final mosaic of the Marciana Marina experiment (not oversampled). (b) Orthophotomap of the Marciana Marina environment. Note the presence of common features that can be appreciated in both representations (indicated by red arrows).

distributed along the trajectory; however, since the estimated trajectory and the GPS track are fixed with respect to the first position, the error might seem larger at the end. Notice also that the discrepancy is around 15 m, which is barely 0.7% of the total trajectory. The acoustic mosaic built

using the estimated global positions [shown in Figure 21(a)] presents an overall view of the surveyed area with a continuous and uniform appearance. A result of this type is of special interest not only to observe the harbor features and their spatial arrangement but also because it enables

us to perceive features that otherwise would be difficult to distinguish given the low resolution and SNR of the acquisition sonar.

By georeferencing the mosaic, we can compare it with an orthophotomap of the harbor environment where the sonar data were gathered [Figure 21(b)] and correlate the presence of scene features (isolated rocks in the left part of the image) in both representations.

## 7. CONCLUSIONS

This paper has described a registration methodology for aligning FLS images, proving its utility in mapping underwater environments. The pairwise registration of sonar frames has been solved by a Fourier-based method that, unlike feature-based methods, takes into account all the image information in the alignment process. We have tailored the phase correlation technique for the special case of FLS image alignment and have compared its performance with an existing region-based FLS registration method. Our approach has shown superior performance in the alignment of both consecutive and nonconsecutive frames, proving to be robust and accurate despite the complicated nature of the sonar images. The main restriction of the method is that it can only register images that differ in small orientations. Although this is not an issue in consecutive or near-consecutive frame registrations, it should be taken into account to obtain successful loop closures.

The registration method has been integrated in a mosaicing pipeline with global alignment being performed by means of a pose-based graph optimization. The high frame rate of the sonar is exploited to achieve local robustness in the initial guess of the graph, and loop closure situations are detected and incorporated to enforce global consistency.

Reported results include mosaics of different characteristics, ranging from small-sized objects to large natural environments, gathered with different sonar models and illustrated with application environments typically affected by low-visibility conditions.

Due to the increase in SNR and resolution with respect to individual frames, the rendered mosaics can serve as an enhanced basis to perform subsequent processing in other applications such as object recognition or terrain classification. In a parallel work, we have started to exploit the advantage of mosaiced frames to apply pattern matching with higher reliability than by using single frames (Hurtós, Palomeras, & Salvi, 2013c).

It is worth pointing out that although the present work focuses on the offline generation of acoustic mosaics, the proposed registration method has the potential of running in real time. Future work will concentrate on integrating the method in an online SLAM framework to help constrain the dead-reckoning drift and thus enable long-term autonomous operations in turbid environments.

## ACKNOWLEDGMENTS

This work has been supported by the FP7-ICT-2011-7 project PANDORA-Persistent Autonomy through Learning, Adaptation, Observation and Re-planning (Ref. 288273) funded by the European Commission, and the Spanish Project AN-DREA/RAIMON (Ref. CTM2011-29691-C02-02) funded by the Ministry of Science and Innovation. The authors would like to thank Soundmetrics Corp., Bluefin Robotics Corp. and the Centre for Maritime Research and Experimentation (CMRE) for providing some of the sonar data.

## REFERENCES

- ANFFT Package. (2013). Retrieved December 31, 2013, from <https://code.google.com/p/anfft>.
- Aykin, M., & Negahdaripour, S. (2012). On feature extraction and region matching for forward scan sonar imaging. *Oceans*, 1–9.
- Aykin, M. D., & Negahdaripour, S. (2013). On feature matching and image registration for two-dimensional forward-scan sonar imaging. *Journal of Field Robotics*, 30(4), 602–623.
- Balci, M., & Foroosh, H. (2006). Subpixel estimation of shifts directly in the Fourier domain. *IEEE Transactions on Image Processing*, 15(7), 1965–1972.
- Baumgartner, L. J., & Wales, N. S. (2006). Assessment of a dual-frequency identification sonar (DIDSON) for application in fish migration studies. NSW Department of Primary Industries.
- Biber, P., & Straßer, W. (2003). The normal distributions transform: A new approach to laser scan matching. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, 2003 (IROS 2003)* (vol. 3, pp. 2743–2748). IEEE.
- Bluefin Robotics, Corp. (2013). Retrieved December 31, 2013, from <http://www.bluefinrobotics.com>.
- BlueView Technologies, Inc. (2013). Retrieved December 31, 2013, from <http://www.blueview.com>.
- Bülow, H., & Birk, A. (2011). Spectral registration of volume data for 6-DOF spatial transformations plus scale. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)* (pp. 3078–3083).
- Bülow, H., Birk, A., & Unnithan, V. (2009). Online generation of an underwater photo map with improved Fourier Mellin based registration. *OCEANS 2009–EUROPE*, 1–6.
- Bülow, H., Pflingstorn, M., & Birk, A. (2010). Using robust spectral registration for scan matching of sonar range data. In *7th Symposium on Intelligent Autonomous Vehicles (IAV)*. IFAC.
- Chailloux, C. (2005). Region of interest on SONAR image for non symbolic registration. In *Proceedings MTS/IEEE OCEANS* (pp. 1–5).
- Chen, Q.-S., Defrise, M., & Deconinck, F. (1994). Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(12), 1156–1168.

- Costello, C. (2008). Multi-reference frame image registration for rotation, translation, and scale. Technical report, DTIC Document.
- De Castro, E., & Morandi, C. (1987). Registration of translated and rotated images using finite Fourier transforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5, 700–703.
- Elibol, A., Gracias, N., Garcia, R., Gleason, A., Gintert, B., Lirman, D., & Reid, R. P. (2011). Efficient autonomous image mosaicing with applications to coral reef monitoring. In *Proceedings of the Workshop on Robotics for Environmental Monitoring held at IEEE/RSJ IROS, San Francisco* (pp. 50–57).
- Eustice, R., Pizarro, O., Singh, H., & Howland, J. (2002). UWIT: Underwater image toolbox for optical image processing and mosaicing in MATLAB. In *Proceedings of the International Underwater Technology Symposium* (pp. 141–145).
- Eustice, R. M., Pizarro, O., & Singh, H. (2008). Visually augmented navigation for autonomous underwater vehicles. *IEEE Journal of Oceanic Engineering*, 33(2), 103–122.
- Fairfield, N., Jonak, D., Kantor, G. A., & Wettergreen, D. (2007). Field results of the control, navigation, and mapping systems of a hovering AUV. In *Proceedings of the 15th International Symposium on Unmanned Untethered Submersible Technology*, Durham, NH, USA.
- Fischler, M., & Bolles, R. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381–395.
- Foroosh, H., Zerubia, J. B., & Berthod, M. (2002). Extension of phase correlation to subpixel registration. *IEEE Transactions on Image Processing*, 11(3), 188–200.
- Galceran, E., Djapic, V., Carreras, M., & Williams, D. P. (2012). A real-time underwater object detection algorithm for multi-beam forward looking sonar. In *IFAC's workshop on Navigation, Guidance and Control of Underwater Vehicles (NGCUV)*, Porto, Portugal.
- Gracias, N. R., Van Der Zwaan, S., Bernardino, A., & Santos-Victor, J. (2003). Mosaic-based navigation for autonomous underwater vehicles. *IEEE Journal of Oceanic Engineering*, 28(4), 609–624.
- Hoge, W. S. (2003). A subspace identification extension to the phase correlation method [MRI application]. *IEEE Transactions on Medical Imaging*, 22(2), 277–280.
- Hover, F. S., Eustice, R. M., Kim, A., Englot, B., Johannsson, H., Kaess, M., & Leonard, J. J. (2012). Advanced perception, navigation and planning for autonomous in-water ship hull inspection. *The International Journal of Robotics Research*, 31(12), 1445–1464.
- Hurtós, N., Cufi, X., Petillot, Y., & Salvi, J. (2012). Fourier-based registrations for two-dimensional forward-looking sonar image mosaicing. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2012)* (pp. 5298–5305). IEEE.
- Hurtós, N., Cufi, X., & Salvi, J. (2013a). A novel blending technique for two dimensional forward looking sonar mosaicing. In *IEEE/MTS OCEANS 2013, San Diego*.
- Hurtós, N., Cufi, X., & Salvi, J. (2014). Rotation estimation for two-dimensional forward-looking sonar mosaicing. In *Armada, M. A., Sanfeliu, A., & Ferre, M. (eds.), ROBOT2013: First Iberian Robotics Conference, vol. 252 of Advances in Intelligent Systems and Computing* (pp. 69–84). Springer International Publishing.
- Hurtós, N., Nagappa, S., Cufi, X., Petillot, Y., & Salvi, J. (2013b). Evaluation of registration methods on two-dimensional forward-looking sonar imagery. In *MTS/IEEE Oceans 2013, Bergen*.
- Hurtós, N., Palomeras, N., & Salvi, J. (2013c). Automatic detection of underwater chain links using a forward-looking sonar. In *MTS/IEEE Oceans 2013, Bergen*.
- Johannsson, H., Kaess, M., Englot, B., Hover, F., & Leonard, J. (2010). Imaging sonar-aided navigation for autonomous underwater harbor surveillance. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, (pp. 4396–4403).
- Kaess, M., Ranganathan, A., & Dellaert, F. (2008). iSAM: Incremental smoothing and mapping. *IEEE Transactions on Robotics*, 24(6), 1365–1378.
- Keller, Y., Shkolnisky, Y., & Averbuch, A. (2005). The angular difference function and its application to image registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(6), 969–976.
- Kim, K., Intrator, N., & Neretti, N. (2004). Image registration and mosaicing of noisy acoustic camera images. In *Proceedings of the 11th IEEE International Conference on Electronics, Circuits and Systems, ICECS 2004* (pp. 527–530).
- Kim, K., Neretti, N., & Intrator, N. (2005). Mosaicing of acoustic camera images. *IEEE Proceedings—Radar, Sonar and Navigation*, 152(4), 263–270.
- Kinsey, J., Eustice, R., & Whitcomb, L. (2006). A survey of underwater vehicle navigation: Recent advances and new challenges. In *Proceedings of the 7th IFAC Conference on Manoeuvring and Control of Marine Crafts, Lisbon, Portugal*.
- Kummerle, R., Grisetti, G., Strasdat, H., Konolige, K., & Burgard, W. (2011). g2o: A general framework for graph optimization. In *Proceedings of the IEEE International Robotics and Automation (ICRA) Conference* (pp. 3607–3613).
- Leonard, J., Bennett, A., Smith, C., & Feder, H. (1998). Autonomous underwater vehicle navigation. Technical memorandum 98-1. MIT Marine Robotics Laboratory.
- Li, L., Qu, Z., Zeng, Q., & Meng, F. (2007). A novel approach to image roto-translation estimation. In *IEEE International Conference on Automation and Logistics* (pp. 2612–2616).
- Lirman, D., Gracias, N., Gintert, B., Gleason, A., Deangelo, G., Dick, M., Martinez, E., & Reid, R. (2010). Damage and recovery assessment of vessel grounding injuries on coral reef habitats by use of georeferenced landscape video mosaics. *Limnology and Oceanography: Methods*, 8, 88–97.

- Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110.
- Lucchese, L., & Cortelazzo, G. M. (2000). A noise-robust frequency domain technique for estimating planar rotations. *IEEE Transactions on Signal Processing*, 48(6), 1769–1786.
- Moré, J. J. (1978). The Levenberg-Marquardt algorithm: Implementation and theory. In *Numerical analysis* (pp. 105–116). Springer.
- Negahdaripour, S. (2012a). On 3-D scene interpretation from FS sonar imagery. In *MTS/IEEE Oceans, 2012* (pp. 1–9).
- Negahdaripour, S. (2012b). Visual motion ambiguities of a plane in 2-D FS sonar motion sequences. *Computer Vision and Image Understanding*, 116(6), 754–764.
- Negahdaripour, S., Aykin, M. D., & Sinnarajah, S. (2011). Dynamic scene analysis and mosaicing of benthic habitats by fs sonar imaging—issues and complexities. In *Proceedings of OCEANS 2011* (pp. 1–7).
- Negahdaripour, S., & Firoozfam, P. (2006). An ROV stereovision system for ship-hull inspection. *IEEE Journal of Oceanic Engineering*, 31(3), 551–564.
- Negahdaripour, S., Firoozfam, P., & Sabzmejdani, P. (2005). On processing and registration of forward-scan acoustic video imagery. In *Proceedings of the 2nd Canadian Conference on Computer and Robot Vision* (pp. 452–459).
- Negahdaripour, S., Sekkati, H., & Pirsavash, H. (2009). Optoacoustic stereo imaging: On system calibration and 3-D target reconstruction. *IEEE Transactions on Image Processing*, 18(6), 1203–1214.
- Pfingsthorn, M., Birk, A., Schwertfeger, S., Bülow, H., & Pathak, K. (2010). Maximum likelihood mapping with spectral image registration. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)* (pp. 4282–4287).
- Point Cloud Library (2013). Retrieved December 31, 2013, from <http://www.pointclouds.org>.
- Reddy, B. S., & Chatterji, B. N. (1996). An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Transactions on Image Processing*, 5(8), 1266–1271.
- Ren, J., Jiang, J., & Vlachos, T. (2010). High-accuracy sub-pixel motion estimation from noisy images in Fourier domain. *IEEE Transactions on Image Processing*, 19(5), 1379–1384.
- Ribas, D., Palomeras, N., Ridao, P., Carreras, M., & Mallios, A. (2012). Girona 500 AUV: From survey to intervention. *IEEE/ASME Transactions on Mechatronics*, 17(1), 46–53.
- Roman, C., & Singh, H. (2005). Improved vehicle based multi-beam bathymetry using sub-maps and SLAM. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 3662–3669), Edmonton, Canada.
- Schwertfeger, S., Bülow, H., & Birk, A. (2010). On the effects of sampling resolution in improved Fourier Mellin based registration for underwater mapping. In *7th Symposium on Intelligent Autonomous Vehicles (IAV)*. IFAC.
- Sekkati, H., & Negahdaripour, S. (2007). 3-D motion estimation for positioning from 2-D acoustic video imagery. In *Pattern Recognition and Image Analysis* (pp. 80–88). Springer.
- Smith, R. C., & Cheeseman, P. (1986). On the representation and estimation of spatial uncertainty. *The International Journal of Robotics Research*, 5(4), 56–68.
- Sound Metrics ARIS (2013). Retrieved December 31, 2013, from <http://www.soundmetrics.com/Products/ARIS-Sonars/ARIS-Explorer-3000>.
- Sound Metrics DIDSON (2013). Retrieved December 31, 2013, from <http://www.soundmetrics.com/Products/DIDSON-Sonars>.
- Soundmetrics Corp. (2013). Retrieved December 31, 2013, from <http://www.soundmetrics.com>.
- Tena, I., Reed, S., Petillot, Y., Bell, J., & Lane, D. M. (2003). Concurrent mapping and localisation using side-scan sonar for autonomous navigation. In *Proceedings of the 13th International Symposium on Unmanned Untethered Submersible Technology*, Durham, NH.
- Tritech Gemini (2013). Retrieved December 31, 2013, from <http://www.tritech.co.uk/product/gemini-720i-300m-multibeam-imaging-sonar>.
- Tuytelaars, T., & Mikolajczyk, K. (2008). Local invariant feature detectors: A survey. *Foundations and Trends® in Computer Graphics and Vision*, 3(3), 177–280.
- Vaganay, J., Elkins, M., Willcox, S., Hover, F., Damus, R., Desset, S., Morash, J., & Polidoro, V. (2005). Ship hull inspection by hull-relative navigation and control. In *OCEANS, 2005. Proceedings of MTS/IEEE* (pp. 761–766).
- Vandrish, P., Vardy, A., Walker, D., & Dobre, O. A. (2011). Side-scan sonar image registration for AUV navigation. In *Proceedings of the IEEE Symposium on Underwater Technology (UT) and 2011 Workshop on Scientific Use of Submarine Cables and Related Technologies (SSC)* (pp. 1–7).
- Walter, M. (2008). Sparse Bayesian information filters for localization and mapping. PhD thesis, Massachusetts Institute of Technology.
- Zitova, B., & Flusser, J. (2003). Image registration methods: A survey. *Image and Vision Computing*, 21(11), 977–1000.