

# Fourier-based Registrations for Two-Dimensional Forward-Looking Sonar Image Mosaicing

Natalia Hurtós<sup>1</sup>, Xavier Cuf<sup>1</sup>, Yvan Petillot<sup>2</sup> and Joaquim Salvi<sup>1</sup>

**Abstract**—This paper presents a method to build large-scale mosaics adapted to underwater sonar imagery. By assuming a simplified imaging model, we propose to address the registrations between images using Fourier-based methods which, unlike feature-based methods, prove well suited to handle the characteristics of forward-looking sonar images, such as low resolution, noise, occlusions and moving shadows. The registration between spatially and temporally distant images resulting from loop-closing situations or registrations in featureless areas are feasible, overcoming the main difficulties of feature-based methods. The problem is cast as a pose-based graph optimization, taking into account the uncertainties of the pairwise registrations and being able to incorporate navigation information. After the optimization, a consistent mosaic from different tracklines is generated with increased resolution and higher signal-to-noise ratio than the original images, while the vehicle motion in x,y and heading is also estimated.

## I. INTRODUCTION

A great number of monitoring and mapping tasks in underwater scenarios are carried out in turbid waters and highly-cluttered environments like ship hull and dam inspections or mapping of harbors, lakes and rivers. In these contexts, vehicles equipped only with optical cameras are constrained by their limited visibility range. However, the ongoing development of two-dimensional forward-looking sonars (FLS) opens the possibility of mapping these environments regardless of the visibility conditions.

Similarly to the mosaics built from video frames, a composite view obtained from acoustic images is of high interest since it provides a global overview of the surveyed area and the spatial arrangement of targets or elements to be studied. A key step in mosaicing is the registration of the images according to an underlying transformation model. In video images, the traditional workflow consists on extracting common salient points from two frames and matching them to find the transformation that aligns one to the other.

However, the characteristics of the FLS images pose a challenging problem to the common registration methods used on photo-mosaicing. FLS images have lower resolution and lower signal-to-noise ratio (SNR) with respect to video images. Besides they can be affected by inhomogeneous insonification and intensity alterations due to the change of the sonar's viewpoint.

This work has been partially supported by the FP7-ICT-2011-7 project PANDORA (Ref 288273) funded by the European Commission, the Raimon project CTM2011-29691-C02-02 and the grant BES-2008-006095.

<sup>1</sup>N. Hurtós, X.Cuf and Q.Salvi are with the Computer Vision and Robotics Group (VICOROB), University de Girona, 17071 Girona, Spain {nhurtos, xcuf, qsalvi}@eia.udg.edu

<sup>2</sup>Y. Petillot is with the Ocean Systems Laboratory, Heriot-Watt University, EH14 4AS Edinburgh, U.K. y.r.petillot@hw.ac.uk

Recently, researchers have approached the problem of registering FLS imagery using feature-based methods. In [1] spatially and temporally close image pairs from DID-SON sonar [2] are registered by using a Harris corner detector. Similarly, in [3] Harris features extracted at the third and fourth level of a Gaussian pyramid scale are matched with cross-correlation and used within a mosaicing algorithm. In [4] the complexities of registering FLS images are highlighted showing the difficulty of establishing correct matches using the popular SIFT detector [5]. The difficulties to accurately extract stable features are exacerbated when dealing with spatially or temporally distant sonar images. In this situations, performing a correct alignment employing features at pixel level becomes unfeasible. Consequently, the possibilities of identifying loop closures, which are necessary to enforce global consistency on the final mosaic, are heavily reduced. Other works [6] approach the problem by extracting dense features not at pixel level, but large-scale features obtained from previously segmented and clustered intensity points. This yields more stable features that allow to identify loop closures and use the registration results as constraints in a sonar-aided navigation system. Nevertheless, as all feature-based methods, it requires the presence of highly distinctive features, thus reducing its reliability in featureless areas.

Consequently, it seems natural to explore area-based registration methods which instead of using sparse features make use of the whole information of the images to minimize ambiguities in registration. In this work we propose to take advantage of the Fourier-based methods to register FLS images. The particularities of these methods suggest that they might be appropriate for the registration of FLS imagery since they are, by design, robust to noise, illumination changes and occlusions.

To the best of the authors' knowledge, this is the first work proposing the use of Fourier-based methods to register 2D-FLS images. To enable its applicability, we consider a simplified model of the sonar's imaging geometry under reasonable assumptions and we embed the pairwise registrations in a pose based graph to enforce global consistency. We present results on real data, including mosaics from different vehicle tracklines and corresponding to different motion types (lateral, forward and circular). The mosaics are shown to be consistent even in the presence of highly uniform areas where feature-methods would definitely fail. Furthermore, the presented work shows also the potential use of Fourier-based registration to provide position estimates based on sonar imagery which could be integrated within a sonar-aided navigation framework.

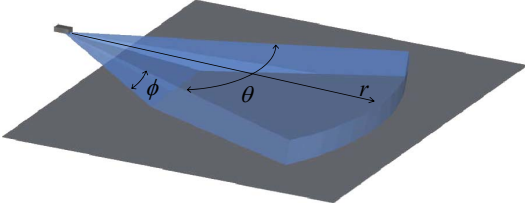


Fig. 1: Imaging sonar geometry ( $r$ : range,  $\theta$ : azimuth,  $\phi$ : elevation).

The rest of the paper is organized as follows. The next section specifies the imaging geometry of the 2D FLS that will be considered throughout the paper. Section 3 reviews the basics of Fourier-based registration methods highlighting its particularities when dealing with FLS images. Section 4 describes the formalization of the problem into a pose-based graph optimization. Mosaics and estimated motion trajectories obtained for different datasets are reported and discussed in section 5. Finally section 6 provides concluding remarks.

## II. IMAGING GEOMETRY OF 2D FORWARD-LOOKING SONARS

In order to register different frames of a FLS it is necessary to understand the image formation process and find out a suitable model to describe the imaging geometry of the sonar. Although the operating frequency, acoustic beamwidth and frame rate varies depending on the particular sonar model, the principle of operation is the same for all of them: the sonar insonifies the scene with an acoustic wave, spanning its field of view in azimuth ( $\theta$ ) and elevation ( $\phi$ ) directions, and the acoustic return is sampled by an array of transducers as a function of range and bearing (Fig. 1). However, given a particular range and bearing it is not possible to disambiguate the elevation angle of the acoustic return since the reflected echo could have originated anywhere along the corresponding elevation arc.

Previous research has analyzed the transformation that relates two acoustic images and has identified the imaging geometry of a FLS. Following the approach of [7], [6] we identify here a simplified model that, though being only an approximation, is suitable to describe the image formation process and to set the basis for the subsequent registration process.

### A. Orthographic Projection Approximation

Without loss of generality, we assume the sonar center at the origin of the world reference frame (Fig. 2). Then, a 3D point with spherical coordinates  $(r, \theta, \phi)$ , where  $r$  corresponds to the sonar's range and  $\theta$  and  $\phi$  to the azimuth and elevation angles respectively, can be defined by the following Cartesian coordinates:

$$\mathbf{x}_w = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} r \cos \theta \cos \phi \\ r \sin \theta \cos \phi \\ r \sin \phi \end{bmatrix} \quad (1)$$

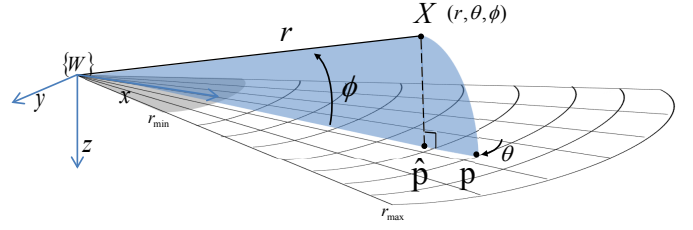


Fig. 2: Sonar projection model.

The projected point  $\mathbf{p}$  in the image plane follows a nonlinear model:

$$\mathbf{p} = \begin{bmatrix} r \cos \theta \\ r \sin \theta \end{bmatrix} \quad (2)$$

The projection described in Eq.2 can be approximated as an orthographic projection onto the image plane. Therefore, the projection  $\mathbf{p}$  of a point  $\mathbf{x}_w$  is approximated by the orthogonal projection  $\hat{\mathbf{p}}$  (Fig. 2). The reader can refer to [7] for the details of the derivation from the nonlinear model to the linear approximation. Essentially it is based on the fact that the nonlinear part is introduced as a function of the elevation angle. Thus, the narrow elevation angle that typically characterizes FLS imposes a tight bound to this non-linear component. Bringing this narrow elevation approximation to the limit yields a linear model in which the sonar can be seen as an orthographic camera. As pointed out by Walter et al., the approximation holds as long as the scene's relief in the elevation direction is negligible compared to the range. It is worth to notice that the imaging geometry under a typical operation scenario falls within this approximation since the sonar device is normally tilted to a small grazing angle to cover a larger portion of the scene.

Hence an orthographic projection model can be defined for the sonar imaging geometry, which can be seen as an affine transformation followed by a parallel projection on the image plane (Eq. 3). A 3D world point  $\mathbf{x}_w = [x, y, z, 1]^T$ , in homogeneous coordinates, is projected into a point  $\mathbf{u}_I = [u, v, 0, 1]^T$  in the sonar image by applying the orthographic matrix  ${}^I\mathbf{T}_w$ :

$$\mathbf{u}_I \approx \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0_{1 \times 3} & 0 \\ 0_{1 \times 3} & 1 \end{bmatrix} \mathbf{x}_w = {}^I\mathbf{T}_w \mathbf{x}_w, \quad (3)$$

$${}^I\mathbf{T}_w = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

where  $r_{ij}$  are the components of a 3D rotation matrix  $\mathbf{R}$  that defines the rotation of the sonar relative to the world frame and  $t_x$  and  $t_y$  are the  $x$  and  $y$  components of the translation.

Given two different locations of the sonar device, one aligned to the world frame and the other located at an arbitrary transformation  $(\mathbf{R}, \mathbf{t})$ , we can then relate the two

sonar image projections  $\mathbf{I}_1$  and  $\mathbf{I}_2$  by:

$$\mathbf{I}_2 \mathbf{T}_{\mathbf{I}_1} = \mathbf{I}_2 \mathbf{T}_{\mathbf{w}} (\mathbf{I}_1 \mathbf{T}_{\mathbf{w}})^{-1} = \begin{bmatrix} r_{11} & r_{12} & 0 & t_x \\ r_{22} & r_{23} & 0 & t_y \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5)$$

Since the  $z$  coordinate of the projections lies on the image plane, we can simplify  $\mathbf{I}_2 \mathbf{T}_{\mathbf{I}_1}$  and obtain the relation that maps the points from one sonar image to the other.

$$\begin{bmatrix} u_{\mathbf{I}_2} \\ v_{\mathbf{I}_2} \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & t_x \\ r_{22} & r_{23} & t_y \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} u_{\mathbf{I}_1} \\ v_{\mathbf{I}_1} \\ 1 \end{bmatrix} \quad (6)$$

Johannsson et al. present in [6] a study on how this image formation model affects the accuracy of the registration of two frames. Basically, under a translation of the sonar the error introduced by the orthographic projection approximation is a function of the distance in the  $x,y$  plane and the vertical distance to the point. Under a suitable imaging configuration this error is normally of few cm, and decreases as the targets are situated further from the sonar's origin. On the other hand, the projection preserves the change in azimuth angles so if the sonar rotates respect to its vertical axis ( $z$ ) the projection on the sonar images 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 the reflected intensities but does not introduce a change on the projection of the points. On the other hand, changes in roll would introduce an error as deviating from the horizontal planar assumption, but we can consider it negligible due to the usual stability of the underwater vehicles in this degree of freedom.

Hence, under the aforementioned assumptions and ensuring a proper imaging configuration, the geometry model relating two sonar views can be simplified down to a Euclidean transformation, which is comprised inside the scope of transformations that Fourier-based methods can resolve. By considering this model, the determination of the translation and rotation parameters between two sonar images provides the transformation to register them, but also gives a direct estimate of the sonar translation in  $x,y$  directions and its heading angle. This holds under our assumptions except when the sonar changes the tilt angle between images. In that case, the estimation of the  $x,y$  position can not be directly related, but still can be estimated if the perpendicular distance of the sonar to the reference plane is known. The ability to estimate the vehicle heading and  $x,y$  position from the sonar imagery is of particular interest in the context of sonar-aided navigation since those are typically the measurements that suffer from bias or drift, while the depth, pitch and roll measurements can be more reliable estimated from pressure and gravity sensors.

### III. FOURIER-BASED REGISTRATION

Fourier-based methods, and in particular the well-known phase correlation algorithm [8][9] that will be explained hereafter, have been employed successfully in several image

processing tasks such as image registration, pattern recognition, motion compensation or 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 [10][11]. However, when dealing with video images, feature-based methods are generally more popular since their high resolution and SNR allow to easily extract stable features 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 [12][13] and some other works have employed it in the registration of 2D or 3D sonar range scans [14][15].

#### A. Phase Correlation Method

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  $(x_0, y_0)$ , namely

$$f(x,y) = g(x - x_0, y - y_0) \quad (7)$$

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

$$F(u,v) = G(u,v) e^{-i(ux_0 + vy_0)} \quad (8)$$

Their normalized cross power spectrum is given by

$$C(u,v) = \frac{F(u,v)G^*(u,v)}{|F(u,v)G^*(u,v)|} = e^{-i(ux_0 + vy_0)} \quad (9)$$

where  $*$  is the complex conjugate. The normalizing denominator in this equation is equivalent to a pre-whitening 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. These effects are indeed present in FLS images which are commonly affected by inhomogeneous intensity patterns due to different sensitivity of the lens or transducers according to their position in the sonar's field of view [1]. The most common way to solve Eq. 9 for  $(x_0, y_0)$  is to apply the inverse Fourier transform to  $C(u,v)$ , which will give the so-called phase correlation matrix as  $PCM(x,y)$ . In the ideal case, this corresponds to a 2D impulse (Dirac function) centered at  $(x_0, y_0)$ , that directly leads to the identification of the integer displacements (Fig. 3a). In the presence of noise or other perturbations in the images the phase correlation matrix will be noisier but as long as it contains a dominant peak the offsets can be retrieved.

The method can be extended to estimate the rotation and scaling parameters between two images [16][9] by applying

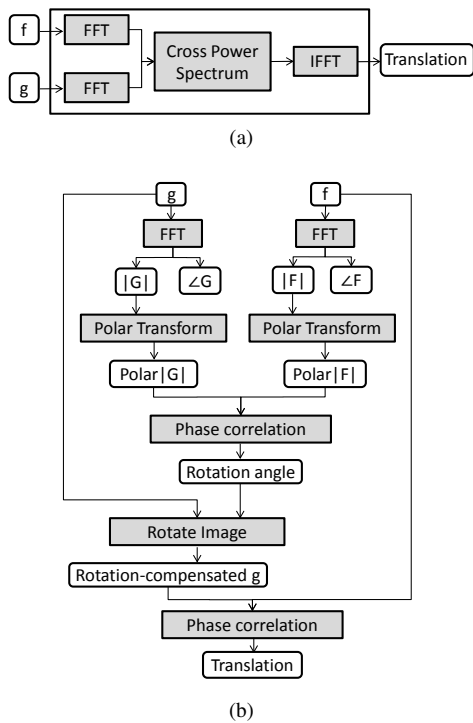


Fig. 3: (a) Basic phase correlation workflow to detect translational shifts. (b) Workflow used for the estimation of rotations and translations. The rotation is estimated separately as a linear shift in the polar transform of the FT magnitude. Then, phase correlation is applied again with the rotation-compensated image to retrieve the translational offsets.

the same procedure to the log-polar magnitude of the Fourier Transforms of the images. In the log-polar coordinates, the differences of the images in rotation and scale are represented by linear shifts. According to the properties of the Fourier transform, its magnitude is invariant to the translation so, the rotation and scaling parameters can be determined in a manner invariant to translation. In our case, since the sonar images do not present the scale ambiguity, it is only necessary to determine the rotation parameter. The full algorithm flow is illustrated in the diagram of Fig. 3b.

Out of this phase correlation principle we can find numerous algorithm variations in the literature. Most of them are concerned with finding sub-pixel displacements after the maximum correlation peak has been determined in the spatial domain within integer accuracy [17][18]. Other attempt the use of high order statistics such as the bispectrum instead of the power spectrum to achieve higher resiliency to noise [19]. Some authors propose also variations on the way to detect rotations without the use of polar transforms [20]. Finally, a substantially different kind of approaches try to recover the offsets by working only in the frequency domain by looking at the shifts as the slopes of a plane fit to the phase difference data [21]. Although we have implemented the most relevant of these methods and we are currently evaluating them when dealing with FLS images, here we are interested in verifying

the general applicability of the Fourier-based methods in FLS image registration. Therefore, from now on we will refer to the standard phase correlation algorithm.

### B. Accuracy of Phase Correlation for FLS Image Registration

The accuracy of the phase correlation algorithm is highly influenced by the nature of the FLS images. Its inherent low SNR, non-uniform resolution in Cartesian space, intensity changes due to different viewpoints or shadow variations result in noise that is introduced as arbitrary peaks in the phase correlation matrix, reducing the amplitude of the true registration peak. While a more detailed study of these factors and their particular effect would be of interest, what we want to verify is that despite their presence the method is able to latch to the correct registration point.

Additionally, there are a series of factors, not linked to the image nature itself that affect also the accuracy of the registration under Fourier-based methods. The most critical one is the aliasing due to the edge effects. The phase correlation theory described previously holds for periodic signals and 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 that can cause the so-called boundary effects. The abrupt transitions generated between the edges when the images are tiled result in high-frequency components appearing in the Fourier spectrum which may also be aliased to lower frequency components altering substantially the phase correlation matrix. In a similar manner, the boundaries generated by the non-rectangular field of view of the FLS images in Cartesian coordinates can introduce a rotationally independent aliasing. According to the Fourier properties, a rotation in the spatial domain rotates the corresponding Fourier spectrum by the same amount. However, the high-frequency components of the image boundaries do not depend on the image content and do not rotate, introducing a strong false peak around the  $0^\circ$  rotation that can hide the location of the true peak. In order to minimize these effects it is typical to perform a windowing operation before the FFT is performed. In our case, a mask that tapers the boundaries of the FLS images in Cartesian coordinates is applied to the images prior to any FFT computation and a Hamming window is applied to the magnitude spectra before computing again the FFT to estimate the rotation.

Other factors that also impact the registration accuracy are the interpolation errors introduced by the polar transform of the FT magnitude or the errors derived from the approximation to the Euclidean model. The content of non overlapping areas in the image is also an error source. However it is worth to highlight the robustness of the phase correlation method when registering images with low-overlap content. This is explained by the fact that the phase correlation matrix is a 2D discrete Dirichlet function whose distribution is highly non-Gaussian, whereas the noise process due to non-overlapping areas is highly Gaussian [22]. It is thus possible to identify the signal even in high noise power, which enables

the registration of a high number of non-consecutive image pairs.

In order to evaluate quantitatively the registration accuracy of the phase correlation method we would need ground truth data of sonar frame-to-frame registrations. Without it, we adopt an evaluation scheme similar to [23]. From a large dataset of FLS images acquired during a vehicle trajectory we construct smaller closed trajectories by picking overlapping images along similar path lengths, in a way that the first and last image coincide. Hence, the comparison between the first and last position and angle can be used to assess the accuracy of the registration method. To this end, we compute the errors of 20 trajectories of about 10 images describing an arbitrary path of approximately 3 meters each. The average image rotation and translation errors are less than  $0.7^\circ$  and 2 pixels, which according to the resolution of our Cartesian sonar images is equivalent to 2.5cm.

#### IV. POSE-BASED GRAPH OPTIMIZATION

The registration method described so far is intended to compute the relative transformation between pairs of overlapping images. In order 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 non-consecutive views is obtained. However, it is well-known that chaining transformations over long sequences is prone to cumulative error [24]. With the aim of obtaining a globally consistent set of transformations to compose a mosaic, the problem is reshaped into a pose based graph optimization. We formulate a least squares minimization problem to estimate the maximum likelihood configuration of the sonar images based on the pairwise constraints between consecutive and non-consecutive frames. As our main interest here is the mosaic generation, we adopt an off-line approach. However if the registration constraints were to be used in a motion estimation framework, they could be integrated within efficient on-line methods that have been developed to optimize pose-based graphs, such as incremental smoothing and mapping (iSAM) [25].

##### A. Graph Definition

Each vertex of the graph represents the position of the sonar image on the mosaic plane and the registration constraints are represented as edges linking the two correspondent vertices. Let  $\mathbf{x} = (x_1, \dots, x_n)^T$  be a vector of parameters, where  $\mathbf{x}_i = (x_i, y_i, \theta_i)$  describes the position (in pixels) and orientation of the vertex  $i$ . The initial positions of the vertices can be estimated using the chained transformations between consecutive image pairs. Let  $\mathbf{z}_{i,j}$  and  $\Omega_{i,j}^z$  be respectively the mean and information matrix of the transformation of node  $i$  from node  $j$  obtained from the registration algorithm on the image pair  $(i, j)$ .

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

$$\mathbf{e}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{z}_{ij}) = \mathbf{z}_{ij} \ominus (\mathbf{x}_j \ominus \mathbf{x}_i) \quad (10)$$

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

$$\mathbf{x}_j \ominus \mathbf{x}_i = \begin{bmatrix} (x_j - x_i)\cos\theta_i + (y_j - y_i)\sin\theta_i \\ -(x_j - x_i)\sin\theta_i + (y_j - y_i)\cos\theta_i \\ \text{normAngle}(\theta_i - \theta_j) \end{bmatrix} \quad (11)$$

Essentially the error function measures how well the position blocks  $\mathbf{x}_i, \mathbf{x}_j$  satisfy the constraint  $\mathbf{z}_{ij}$ . The problem is cast in a least-squares minimization of the error function taking into account the information matrices as shown in [26]. If a good initial guess of the parameters is known, Levenberg-Marquardt algorithm [27] can be used to obtain a numerical solution. Therefore, if enough constraints can be extracted using pairwise registrations we can reach a solution relying solely on the sonar images. However, navigation information is typically available and it can be employed. Assuming the model presented in section II the translations and heading of the sonar map into translations and rotations of the projected points in the images. Thus, the navigation increments between the different vehicle poses can be converted into pixels by using the resolution of the sonar images and used to initialize the graph vertices  $\mathbf{x}$ . If the navigation data is accurate, this will give a better starting point for the optimization rather than the poses obtained from chaining the consecutive image transformations.

##### B. Discrimination of Successful Registrations

A measure is required to discriminate whether a registration is successful or not, to avoid introducing false constraints (e.g. in the case of two images that are not overlapping at all). Recalling the description of the method in section III, the amplitude of the peak can be used as a direct measure of the degree of congruence between the two images. Intuitively, we need a measure to quantify how much the peak stands from the rest of the data in the phase correlation matrix. To this end we normalize the phase correlation matrix and compute a peak-to-noise ratio (PNR) [28] through the following equation:

$$PNR = \frac{A_{peak}}{1 - A_{peak}}, A_{peak} = \max_{x,y}(PCM(x, y)) \quad (12)$$

Fig. 4 shows the PNR measure while we attempt to register a frame with progressively more distant frames. The last correctly registered frame (number 45) has translated 91 pixels with respect to the reference one, which in this particular case and given the non-rectangular FOV of the images correspond to a 32% of overlap. After this point the images fail to register and it can be observed that the PNR measure remains small. Hence, setting a threshold on the measure above that value allows to discriminate between successful or failed registrations.

According to the PNR value, we will only include the valid registrations in the graph. However, given the high refresh rate of the FLS sonars (typically around 10Hz), it would be highly computationally expensive to attempt the registration of each frame to all the rest of a sequence's frames. Hence, for each image we only try to register those frames that have been acquired within a pre-established range.

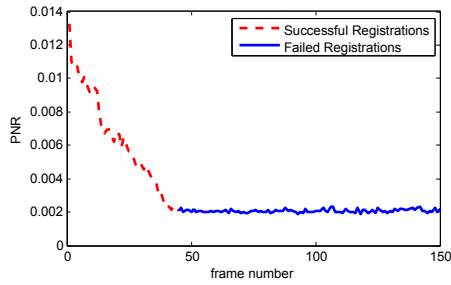


Fig. 4: PNR measure along the registration of progressively distant frames.

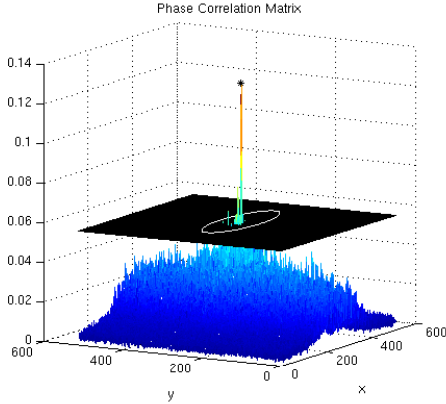


Fig. 5: Determination of the uncertainty of a pairwise registration from the phase correlation matrix values.

### C. Determination of the Registration Uncertainty

For those pairs of images that are actually overlapping it is required to define an uncertainty measure to quantify how trustworthy is their registration. The amplitude and extent of peak values located in the surroundings of the main peak account for localization inaccuracies of the registration. Hence, at a given amplitude the phase correlation surface is cut and we extract the variances in  $x$  and  $y$  directions from the contributing peaks above the cut, as shown in Fig. 5. The cut height has been empirically set to half the amplitude of the main peak. The obtained variance values are afterwards inverted and used to fill in the information matrix  $\Omega^z$  of our pose-graph formulation.

## V. RESULTS

Several experiments have been carried out on real data to evaluate the performance of the method.

### A. Ship Hull Inspection

The first experiment uses the data of a ship hull inspection performed with a DIDSON sonar. The vehicle executed a trajectory of several tracklines across the side of the hull collecting a total of 4420 sonar images during 45 minutes. The tilt of the sonar is changed along the experiment so as to adapt the images to the curvature of the hull and facilitate better conditions for the planar assumption. Since the vehicle is moving basically in surge and heave degrees of freedom,

the model under consideration is an Euclidean transformation with only the two translational components. Small deviations from this model can be corrected by the global optimization, by allowing it to accommodate for small rotations. Fig. 6 shows the result of the mosaic built after the optimization.

To reduce computational costs and unnecessary computations we have used one out of every five images of the full dataset, that have been attempted to register pairwise with its neighbor images inside a  $2m$  distance radius. Hence, about 850 frames from different tracklines have been successfully registered providing a visually coherent image of the hull which supports the applicability of the method. For visualization purposes the inhomogeneous insonification pattern present in the images has been identified by averaging all the dataset frames and subsequently subtracted from all of them. Additionally, a blending process based on averaging the intensities of the overlapping pixels has been applied to generate a seamless composite view.

The average runtime per registration under a MATLAB environment is 0.12 seconds using a Intel Core Duo 3.4Mhz, thus taking about 1 hour to compute the full set of links (up to 23214) for the whole dataset. The optimization, carried out using the General Graph Optimization (g2o) framework [26] has converged with less than 40 iterations taking a total of 2.6 seconds. In this case, no navigation information has been used to set the initial locations of the vertices, thus relying solely on the information extracted from the registrations. However, when navigation is used, the optimization leads to a similar solution.

The vehicle position, which we consider here centered at the sonar's location, can be derived from the result of the optimized image poses. Since the sonar has undergone changes in tilt, its  $x, y$  position is not directly derived from the resulting image poses, but it can be estimated under our planar assumption since we know the range values and the distance from the vehicle to the hull. Since no ground truth is available, a comparison of the resulting motion estimation is performed with the trajectory derived from the navigation sensors on-board (Fig. 6b). As it is well-known, the trajectory computed from the navigation data might suffer also from errors and drift, but all in all and in absence of ground truth it is still a good reference. As can be seen, the two trajectories are closely similar, which again testifies the accuracy of the method. Actually, if we map the images over the navigation positions, several mismatches occur whereas using the resulting optimized poses the mosaic presents a higher coherence which suggest that our method is able to correct the small errors in the navigation.

### B. Mosaicing a Featureless Area

The purpose of this experiment is to prove the ability of the method when dealing with featureless areas. In the employed dataset the vehicle performs a trajectory passing over some circular grids, those being the only prominent features in all the vehicle's path. Fig. 7a shows the final mosaic, built from one out of every five images of the total 2100 frames of the



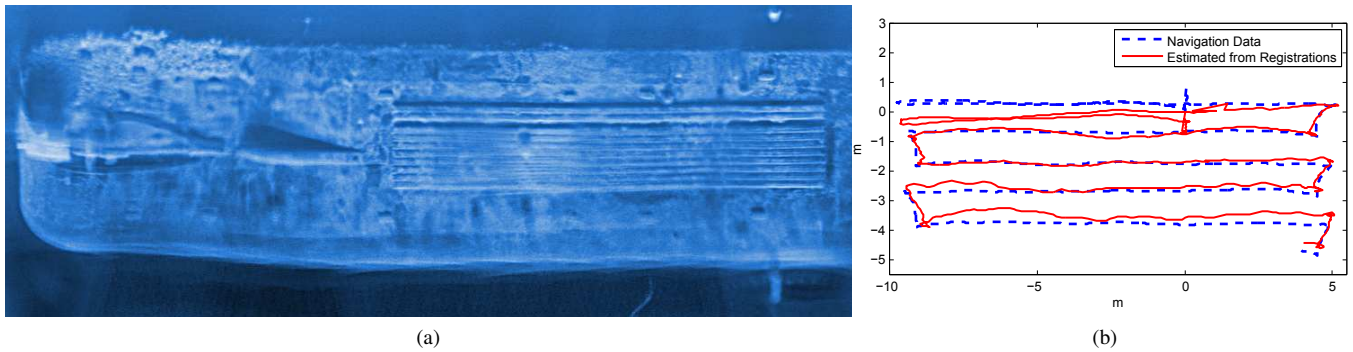


Fig. 6: (a) Mosaic of the ship hull comprising 844 frames. (b) Comparison between the navigation data and the estimated trajectory.

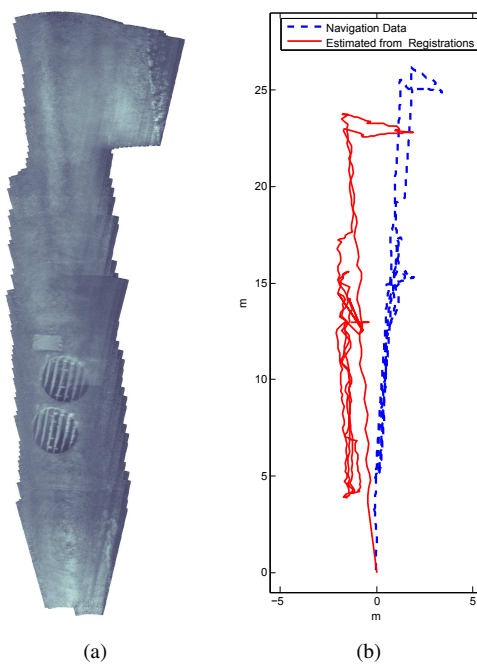


Fig. 7: (a) Mosaic of a featureless dataset with 388 frames. (b) Comparison between the navigation data and the estimated trajectory.

full dataset. No navigation information has been used to set the initial locations of the vertices.

Due to the lack of salient features, one could argue that it is difficult to know from the mosaic whether the images are correctly registered or not. However, the estimated motion from the registrations gives again a coherent trajectory compared to the navigation data, as illustrated in Fig. 7b. Although some drift is present, which could possibly be due to the dead-reckoning, the main trend of the estimated motion is in agreement with the navigation data. This shows the potential of the method to perform mapping in turbid environments such as rivers, lakes or harbors, characterized by the lack of distinguishable objects and prominent shapes.

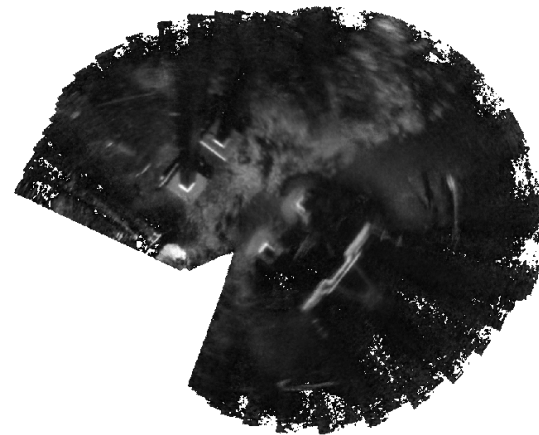


Fig. 8: Mosaic under a rotational movement

### C. Mosaic Under Rotations

Although the common vehicle trajectories when performing mapping consist in straight trajectories we want to show also the ability of the method to register FLS images under rotational movements. Unfortunately, we do not have a large dataset where the FLS images are affected from big rotational movements. However, we used a small dataset from BlueView Technologies [29] where the sonar undergoes a rotation around some bridge footers. Fig. 8 shows the generated mosaic from 147 frames. Although small, the mosaic proves the method's capability to align the frames giving a composite view from the bridge footers.

### D. Super-resolution

The outcome of averaging the overlapping sonar images yields to a denoising of the final mosaic, achieving an improvement in terms of SNR compared to a single image frame (see Fig. 9). However, since the obtained transformations lead to a subpixel alignment of the images, we can perform super-resolution and achieve also improvements in terms of the resolution of the final mosaic. This is of special interest given that a pixel on the sonar's sampling space (range, bearing) maps to a collection of pixels with the same

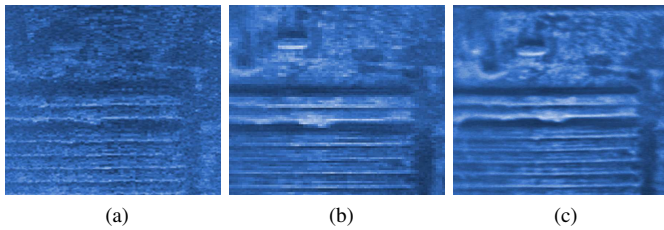


Fig. 9: (a) Single frame (b) Mosaic at the same resolution as the individual frames (c) Mosaic four times the resolution of the individual frames.

intensity in the Cartesian frame, resulting in a nonuniform resolution that degrades the visual appearance. Therefore, making a finer grid and applying the subpixel transformations to the images leads to an overall enhancement. Fig. 9c shows a portion of a mosaic generated with four times the original image resolution where the improvement on the details can be clearly appreciated.

## VI. CONCLUSIONS

The registration of FLS images poses a serious challenge onto the traditional feature-based registration techniques commonly employed on video images. Under reasonable assumptions the imaging geometry of the FLS has been simplified down to an Euclidean transformation thus enabling the application of the Fourier-based methods for registration. The phase correlation technique has been applied to the alignment of consecutive and non-consecutive sonar frames proving to be robust and accurate despite the noisy nature of the images and the lack of abundant features.

By integrating the pairwise registrations into a global alignment we are able to generate a consistent composition of the acoustic images. The obtained mosaic offers a global view of the surveyed area while at the same time presents an overall improvement with respect to the individual images in terms of SNR and resolution.

Furthermore, the results support the use of Fourier-based registration in the field of sonar-aided navigation. The method could be integrated within an on-line framework to provide motion estimates from the sonar imagery with the aim of bounding the navigation drift.

## ACKNOWLEDGMENT

The authors would like to thank Seebyte Ltd. and the Office of Naval Research for providing the sonar data.

## REFERENCES

- [1] S. Negahdaripour, P. Firoozfam, and P. Sabzmejdani, "On processing and registration of forward-scan acoustic video imagery," in *Proc. 2nd Canadian Conf. Computer and Robot Vision*, 2005, pp. 452–459.
- [2] "Soundmetrics." [Online]. Available: [www.soundmetrics.com](http://www.soundmetrics.com)
- [3] K. Kim, N. Neretti, and N. Intrator, "Mosaicing of acoustic camera images," *IEEE Proceedings Radar, Sonar and Navigation*, vol. 152, no. 4, pp. 263–270, 2005.
- [4] S. Negahdaripour, M. D. Aykin, and S. Sinnarajah, "Dynamic scene analysis and mosaicing of benthic habitats by FS sonar imaging - issues and complexities," in *Proc. OCEANS 2011*, 2011, pp. 1–7.
- [5] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [6] H. Johansson, M. Kaess, B. Englot, F. Hover, and J. Leonard, "Imaging sonar-aided navigation for autonomous underwater harbor surveillance," in *Proc. IEEE/RSJ Int. Intelligent Robots and Systems (IROS) Conf.*, 2010, pp. 4396–4403.
- [7] M. R. Walter, "Sparse bayesian information filters for localization and mapping," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, 2008.
- [8] E. De Castro and C. Morandi, "Registration of translated and rotated images using finite fourier transforms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 5, pp. 700–703, 1987.
- [9] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Transactions on Image Processing*, vol. 5, no. 8, pp. 1266–1271, 1996.
- [10] R. Eustice, O. Pizarro, H. Singh, and J. Howland, "UWIT: underwater image toolbox for optical image processing and mosaicking in matlab," in *Proc. Int. Underwater Technology Symp.*, 2002, pp. 141–145.
- [11] H. Bulow, A. Birk, and V. Unnithan, "Online generation of an underwater photo map with improved fourier mellin based registration," in *OCEANS 2009 - EUROPE*, May 2009, pp. 1–6.
- [12] C. Chailloux, "Region of interest on sonar image for non symbolic registration," in *Proc. MTS/IEEE OCEANS*, 2005, pp. 1–5.
- [13] P. Vandrish, A. Vardy, D. Walker, and O. A. Dobre, "Side-scan sonar image registration for AUV navigation," in *Proc. IEEE Symp Underwater Technology (UT) and 2011 Workshop Scientific Use of Submarine Cables and Related Technologies (SSC)*, 2011, pp. 1–7.
- [14] H. Blow, M. Pfingsthorn, and A. Birk, "Using robust spectral registration for scan matching of sonar range data," in *7th Symposium on Intelligent Autonomous Vehicles (IAV)*. IFAC, 2010.
- [15] H. Bulow and A. Birk, "Spectral registration of volume data for 6-dof spatial transformations plus scale," in *Proc. IEEE Int. Robotics and Automation (ICRA) Conf.*, 2011, pp. 3078–3083.
- [16] Q.-S. Chen, M. Defrise, and F. Deconinck, "Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 12, pp. 1156–1168, 1994.
- [17] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE Transactions on Image Processing*, vol. 11, no. 3, pp. 188–200, 2002.
- [18] J. Ren, J. Jiang, and T. Vlachos, "High-accuracy sub-pixel motion estimation from noisy images in fourier domain," *IEEE Transactions on Image Processing*, vol. 19, no. 5, pp. 1379–1384, 2010.
- [19] L. Chen and K.-H. Yap, "An effective technique for subpixel image registration under noisy conditions," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 38, no. 4, pp. 881–887, 2008.
- [20] L. Lucchese and G. M. Cortelazzo, "A noise-robust frequency domain technique for estimating planar roto-translations," *IEEE Transactions on Signal Processing*, vol. 48, no. 6, pp. 1769–1786, 2000.
- [21] M. Balci and H. Foroosh, "Subpixel estimation of shifts directly in the fourier domain," *IEEE Transactions on Image Processing*, vol. 15, no. 7, pp. 1965–1972, 2006.
- [22] M. Shah and R. Kumar, *Video registration*. Springer, 2003, vol. 5.
- [23] S. Negahdaripour, H. Assalhi, Y. Petillot, and L. N. Brisson, "Performance and accuracy in visual motion computation from FS sonar video sequences," in *Proc. OCEANS 2010*, 2010, pp. 1–7.
- [24] A. Elibol, N. Gracias, and R. Garcia, "Augmented state-extended kalman filter combined framework for topology estimation in large-area underwater mapping," *Journal of Field Robotics*, vol. 27, no. 5.
- [25] M. Kaess, A. Ranganathan, and F. Dellaert, "iSAM: Incremental smoothing and mapping," *IEEE Trans. on Robotics (TRO)*, vol. 24, no. 6, pp. 1365–1378, Dec. 2008.
- [26] R. Kummerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "G2o: A general framework for graph optimization," in *Proc. IEEE Int. Robotics and Automation (ICRA) Conf.*, 2011, pp. 3607–3613.
- [27] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes*. Cambridge University Press, 1992.
- [28] H. Bulow, A. Birk, and V. Unnithan, "Online generation of an underwater photo map with improved fourier mellin based registration," in *Proc. OCEANS 2009 - EUROPE*, 2009, pp. 1–6.
- [29] "Blueview technologies inc." [Online]. Available: [www.blueview.com](http://www.blueview.com)