Evaluation of Registration Methods on Two-Dimensional Forward-Looking Sonar Imagery

Natàlia Hurtós[1], Sharad Nagappa[1], Xavier Cufí[1], Yvan Petillot[2] and Joaquim Salvi[1]

[1]*Computer Vision and Robotics Group, University of Girona, Spain*

{nhurtos, snagappa, xcufi, qsalvi}@eia.udg.es

[2]*Ocean Systems Laboratory, Heriot-Watt University, U.K*

y.r.petillot@hw.ac.uk

***Abstract*— Two-dimensional forward-looking sonars such as Blueview or DIDSON are becoming a standard sensor in both remotely-operated and autonomous underwater vehicles. Registration of imagery obtained from this sensors is of great interest since it constitutes a key step in several applications like the generation of acoustic mosaics or the extraction of vehicle motion estimates from sonar imagery, specially on poor visibility conditions. However, the characteristics of these sonar images, such as low signal-to-noise ratio, low resolution and intensity alterations due to viewpoint changes pose a challenge to the traditional registration techniques applied on optical images. In this paper, the performance of popular registration methods commonly used in photomosaicing are evaluated on real sonar data, including feature-based methods and an area-based approach. Experiments are carried out on different environments, from man-made structured scenarios to more natural and featureless areas, and under challenging conditions such as viewpoint changes and the presence of different sonar-specific artifacts. After assessing the impact of all these factors on the different registration techniques, we show that Fourier-based registration method stands as the more robust option to register acoustic imagery.**

I. INTRODUCTION

The increasing development of the two-dimensional forward-looking sonars (2D-FLS) which deliver high-resolution acoustic images at near-video framerate is playing a key role in underwater inspection where water visibility does not allow the use of cameras. Inspection on harbor underwater structures, ship hulls, dams or the monitoring of rivers and lakes are some of the applications that can benefit from this growing technology. The capability of seeing in turbid water enables the implementation of tasks commonly performed with vehicle cameras such as image mosaicing of areas of interest or the extraction of motion estimates from the imagery to aid navigation. Nevertheless, all these applications require a non-trivial and crucial step: the registration of the acoustic images. Although image registration is a broadly studied field in other modalities, notably the optical one, it is not yet clear how the particularities of the sonar imaging have an impact on the registration techniques.

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 ANDREA/RAIMON (Ref CTM2011-29691-C02-02) funded by the Ministry of Science and Innovation.

The challenges that the FLS sonar images pose to traditional registration methods have not gone unnoticed by researchers, who in several works conclude that it is necessary to develop a more suitable and specific registration technique to be applied on 2D-FLS images [1][2]. Most of the existing work adopt feature-based approaches traditionally used in optical image registration, although the motivation for choosing one method or the other is not supported by any test. Hence, most of the reported results using feature-based techniques include restrictive examples (i.e few frames, clearly feature-rich environments) and are subject to the registration of spatially and temporally close sonar frames. Recently, some works have pointed out the fact that feature-based techniques at pixel level are not stable on sonar images [3][4] and researchers have proposed the use of techniques based on stable regions [5] or methods that make use of all the image content [6]. However no experimental evidence or numerical comparison is shown among the existing registration methods.

The contribution of this paper is the review and comparison of the most significant registration techniques commonly used on optical images with FLS imagery. Several commonly used registration methods are evaluated on real data, under distinct environments and under challenging conditions such as viewpoint changes. The impact of the sonar's field of view and resolution is also assessed by using imagery of the two main kinds of FLS present in the market today: a DIDSON, from Soundmetrics [7], and a BlueView P900-130 from BlueView Technologies [8]. The results of the comparative experiments allow us to identify which approach performs better for 2D-FLS images and therefore put a first a step towards developing a more suitable registration technique for this type of imagery.

The rest of the paper is organized as follows. The next section provides a background on 2D Forward-Looking imaging sonars including a brief description of its geometry model, an identification of the main challenges encountered when registering FLS data and a summary of related work present in the literature. Section III provides a brief description of several registration techniques whose performance will be analyzed on FLS images. The series of experiments that have been carried out are described in section IV. A discussion on the results is presented in Section V. Finally Section VI

978-1-4799-0002-2/13/$31.00 ©2013 IEEE

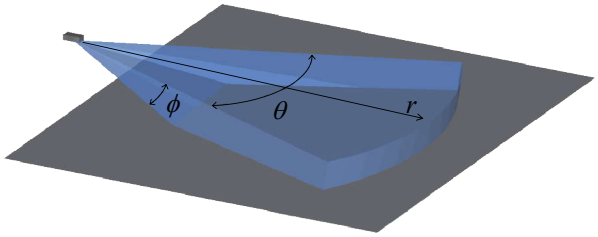


Fig. 1: Imaging sonar geometry (r: range, θ: azimuth, ϕ: elevation).

gives some concluding remarks and points out for future work directions.

II. BACKGROUND

A. Geometry model

Image registration techniques are chosen in accordance with an underlying transformation model. In optical cameras the most general case is to relate two camera views through a projective transformation, although a model with less degrees of freedom can be used under particular circumstances. To determine the registration model for two FLS views, it is necessary first to understand the image formation process and the involved geometry. The sonar insonifies the scene with an acoustic wave, spanning its field of view in azimuth (θ) and elevation (ϕ) 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.

Hence, under this imaging geometry a projected point in the image plane follows a non-linear model, where the nonlinear part is introduced as a function of the elevation angle. Given the narrow elevation angle that FLS typically have (i.e 6° for DIDSON and 10° for BlueView) this non-linear component is bounded. Bringing this narrow elevation approximation to the limit yields a linear model in which the sonar can be seen as an orthographic camera and can be modeled by an affine transform followed by a parallel projection on the image plane [4]. As pointed out by Walter et al. [4], the approximation holds as long as the scene's relief in the elevation direction is negligible compared to the range. It is worth noting that in common vehicle operations this is a reasonable assumption since the sonar is normally tilted to a small grazing angle to cover a large portion of the scene. Hence, under suitable imaging conditions and assuming a locally planar environment, the sonar can be considered as an orthographic camera and a point in two views can be related through an Euclidean transformation composed only by rotation and translation [6].

B. Challenges

2D-FLS images exhibit different characteristics that make their registration more challenging for existing techniques.

1) Low resolution: Although they are considered high-resolution sonars, 2D-FLS image resolution is far from the resolution of today's standard cameras which make use of 2D array sensors with millions of pixels. For instance, the DIDSON sonar samples the acoustic returns with an array of 96 transducers of 0.3° beamwidth. The BlueView P900-130 has 768 beams with 1° beamwidth each. Moreover, if it is required to work in Cartesian coordinates, a pixel in the range-bearing space maps to a collection of pixels with the same intensity in the Cartesian frame, resulting in a nonuniform resolution that degrades the visual appearance.

2) Low signal-to-noise ratio (SNR): Similar to other coherent imaging systems such as radar or ultrasound imaging, 2D FLS suffer from low SNR. This is mainly due to the presence of speckle noise introduced by the mutual interference of the sampled acoustic returns.

3) Inhomogeneous insonification: FLS 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. This effect can indeed affect the alignment, causing the registration to latch on those intensity patterns instead of the real image content. However, it can be corrected by means of a preprocessing step that estimates the inhomogeneous illumination pattern from the averaging of a sufficient number of images.

4) Changes in viewpoint: Intensity variations due to a change in the sonar's viewpoint are inherent to the image formation process and are one of the most difficult effects to cope with. Imaging the same scene from two different view points can cause the movement of shadows in the images, occlusions and in general, significant alterations in the visual appearance of the image that complicate the registration process.

5) Other artifacts: Under some circumstances, spurious content can appear in the sonar images causing ambiguity in the registration. It is the case for instance of acoustic returns from the water surface, reverberation artifacts, or cross-talk between sidelobes of different beams that can generate multiple replicas of a target.

C. Previous works on 2D FLS imagery

Since the development of the first FLS, several researchers have tackled the problem of processing FLS imagery and in particular, its registration. In [9] ten sequential frames of DIDSON images are registered by using a Harris corner detector [10] and matched by searching over small local windows. The author reports a ratio of inliers of 50%. Similarly, in [11], Harris features extracted at the fourth and third level of a Gaussian pyramid scale, are also used within a mosaicing algorithm for FLS images. The points are matched by the cross-correlation of small patches surrounding the identified corners. Each frame is registered sequentially with a window of neighboring frames and the reported results show only registration from translational sonar displacements. In [2] the complexities of mosaicing benthic habitats with FLS images are highlighted showing the difficulty of registering DIDSON frames from a natural environment

using the popular SIFT detector [12]. The authors report a very low percentage of inliers in the detection step (12 out of 150) together with the fact that only short displacements could be effectively matched. In [1] some experiments on synthetic data are carried out to assess the accuracy of motion estimation from FLS images. The motion estimation is obtained by establishing the relation with the correspondences in two sonar frames. Different uncertainties (ranging from half centimeter to two centimeters) are assigned to the feature's location in order to study the impact of increasing the sonar's field of view and the number of correspondences. The main conclusion is that a larger field of view improves the accuracy of motion estimates. An experiment with real data is also carried out computing the motion estimation from five different sonar positions, although the method used to detect the features and register them is not specified.

Walter et al. [4] base their system on larger features instead of identifying interest points at pixel level, in the context of a feature-based SLAM for ship hull inspection. They use a complex multistage algorithm that segments the shadows and echoes of different targets by using filters that have been tuned according to the image signatures of discriminating features. Target hypothesis that are consistent with the vehicle motion are output as valid observed features that are later used inside a SLAM framework.

In [5] features based on sharp transitions are extracted from 2D FLS images to aid the navigation in AUV harbour surveillance. The sonar images are first smoothed with a median filter, then the gradient is computed, and the points exceeding a given threshold are finally clustered in the extracted features. The alignment is then performed using the Normal Distribution Transform (NDT) algorithm [13] which adjusts the features in grid cells allowing for some slack so that exact correspondences between points are not required. The successful registrations are integrated in a SLAM framework reporting successful results on different environments.

Finally, our previous work [6] has proposed to explore Fourier-based registration methods which estimate the alignment by the correlation of the image's frequency spectrums. Pairwise registrations are integrated into a global alignment framework to generate consistent acoustic mosaics. Although the work shows the potential of Fourier-based methods for 2D-FLS registration, no comparison with other methods was reported.

III. REGISTRATION TECHNIQUES

The computer vision community has proposed numerous registration methods in the last decades [14] which can be essentially classified in two main groups, feature-based methods and area-based methods. The next sections present some of the most representative techniques which will be evaluated on FLS images.

A. Feature-based methods

Registration with feature-based methods relies on the detection of a limited set of well localized and individually distinguishable points. In certain applications, these features can be lines or specific object shapes. Otherwise, local features can be determined at pixel scale level based on the underlying intensity patterns, which do not have to necessarily correspond to a physical object or a meaningful part of the image.

The traditional pipeline for the feature-based registration of images consists first on 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. Afterwards, there is a matching step where the point-to-point correspondences from the two images are established and finally this information is used to estimate the homography that relates one image to the other, usually by taking into account some outlier rejection scheme such as RANSAC [15].

As most of the feature detectors have been traditionally designed to work on optical images, their performance is not clear on sonar images. There is therefore a need to better understand feature detection on FLS images. An important aspect when choosing a feature detector is its degree of invariance. Feature detectors are invariant to a certain degree of transformations, i.e., they are able to detect the same features even when the image has undergone that particular transformation. In the case of sonar imaging, this transformation can be modeled as an affine one and simplified to an Euclidean transformation (rotation+translation) according to the assumptions reported in Section II. Hence, although detectors have been developed to work under severe viewpoint changes and estimate complex transformations [12][16], the experience of several authors suggest that the best approach is to use the lowest degree of invariance required for the intended application [14] (i.e, the most constrained transformation). Therefore, provided that the estimation of roto-translations is sufficient for our case and the FLS images do not present scale ambiguity, the selected detectors are not necessarily scale or affine invariant. Given the large number of detectors present in the literature, we select a representative one from each basic category based on the kind of structures they target [14].

1) Corners: Corners are points of high curvature. The Harris corner detector [10] targets locations with high variability of the intensity pattern which makes the extracted corners more discriminative and easier to match. The traditional Harris corner detector is only invariant to translation and rotation and stable under light changes, although there are also scale and affine-invariant versions (Harris-Laplace, Harris-Affine).

2) Blobs: Blobs are often treated as complementary features to corner detectors. One of the most popular is the Hessian detector [17] based on derivative expressions, which also has scale and affine invariant versions.

3) Regions: One of the most popular region detectors is the Maximally Stable Extremal Regions (MSER) [18]. This detector extracts stable regions from the image by considering the change in area with respect to the change in intensity of a connected component. Connected components

are defined by thresholding the image at a given gray level. The change of area, normalized by the area of the connected component, is used as the stability criterion. MSER is photometrically and geometrically affine invariant.

For feature extraction and description, the SIFT descriptor [12] has been used. The best candidate match for a feature is found by identifying its nearest neighbor from all the descriptors of the other image. The nearest neighbor is defined as the descriptor with the minimum Euclidean distance from the given descriptor vector.

B. Area-based methods

In contrast to the described feature-based approaches which make use of the sparse feature information, area-based registration methods use all the image content which reduces the ambiguities in the registration. The main drawback is that they cannot handle complex transformations, often being limited to the estimation of Euclidean or similarity transforms. However, using the simplified FLS geometry model that we adopt, the registration of two FLS images falls in its scope of applicability, thus turning the area-based methods into a possible solution for FLS image alignment.

The simplest approach for area-based registration methods is cross-correlation, or template matching. Cross-correlation convolves the images in the spatial domain resulting in high values where the images are best correlated. A common variant is the normalized cross-correlation where the images are first normalized to cope with non-uniform lighting. However, the cross-correlation matrix in the spatial domain often yields broad peaks with poor localization accuracy. In this sense a better technique which has also been used on optical imagery, is phase-correlation or correlation of the image frequency spectrums [19][20].

Phase correlation is a well-known technique that exploits the Fourier shift theorem, which states that a shift between two images is transformed in the Fourier domain into a linear phase shift. Translational displacements are estimated by detecting the main peak in the phase correlation matrix as described in [6]. This can be extended, by working with the Fourier Transform magnitude in the polar domain, to estimate the rotation between the images. As only the phase information is utilized, the method is robust against illumination changes, occlusions and background clutter, thus making it a good candidate technique for the registration of FLS images [6].

IV. EXPERIMENTS

A. Datasets

Three experiments have been carried out with the aim of evaluating the registration methods under different circumstances.

The first experiment aims to explore the impact of the imaged environment on the registration by using two different datasets, one representing data from feature-rich scenarios (i.e manmade environments with regular shapes) and the other seafloor images with general low texture content. More specifically, the first dataset is a DIDSON sequence from a ship-hull inspection where multiple features are detected in each image. The second dataset contains a DIDSON sequence from a seafloor exploration with extremely low texture content except for a few frames where a small grid is observed. Each dataset is composed of 1000 frames and the registration is applied to every 5th frame.

The second experiment aims to assess the effect of the field of view and resolution on the registration procedure. To this end, the first experiment is repeated on two datasets gathered with the BlueView P900-130 device which has a significantly different resolution and field of view as explained in section II-B. Ideally, for this comparison we would need to have the same area imaged with the two devices. However, having imagery from similar kind of environments imaged with each of them (both feature-rich and featureless) is sufficient to recognize the trends and draw some conclusions. Thus, the first dataset comprises images from an underwater dock and the second from a seafloor terrain with some rocks and vegetation. Examples of images used in experiments 1 and 2 are shown in Figure 2.

For the third experiment, several frames have been compiled in two datasets to assess how the different registration techniques deal with the challenges presented by FLS images. The first dataset includes 10 image pairs that insonify the same scene from two different viewpoints. The frames are extracted from loop closing situations (from several different larger datasets), with viewpoint changes as severe as 180°. An example pair of this dataset is shown in figure 3. The second dataset comprises 10 image pairs with outlier content such as multi-path effects or reflections from the surface water.

In all experiments, a mask is applied in order to avoid the detection of features in the image edges, as well as to avoid edge effects in the phase-correlation technique as explained in [6]. Additionally, all frames have undergone a radiometric correction to suppress the inhomogeneous insonification effects via the estimation of a pattern from the average of all sequence frames.

B. Evaluation

In order to evaluate the performance of the different registration techniques, several criteria have been taken into account. Ideally, the estimated data from the different registration techniques should be compared against a ground-truth. In the absence of this, the navigation data from the sequences serves as a good starting point. Several loop-closing situations have been identified and matched manually to be incorporated as constraints in a pose-based graph and thus compensate the drift due to the dead-reckoning. The g2o software [21] has been used to perform a global optimization on the graph composed of the navigation poses and the identified constraints. The estimated final poses are used as a ground-truth for comparison purposes. On the other hand, translation and rotation parameters have been extracted from the computed homographies (in the case of feature-based techniques) or taken directly from the output of the phase-correlation registration. The first criterion of evalua-

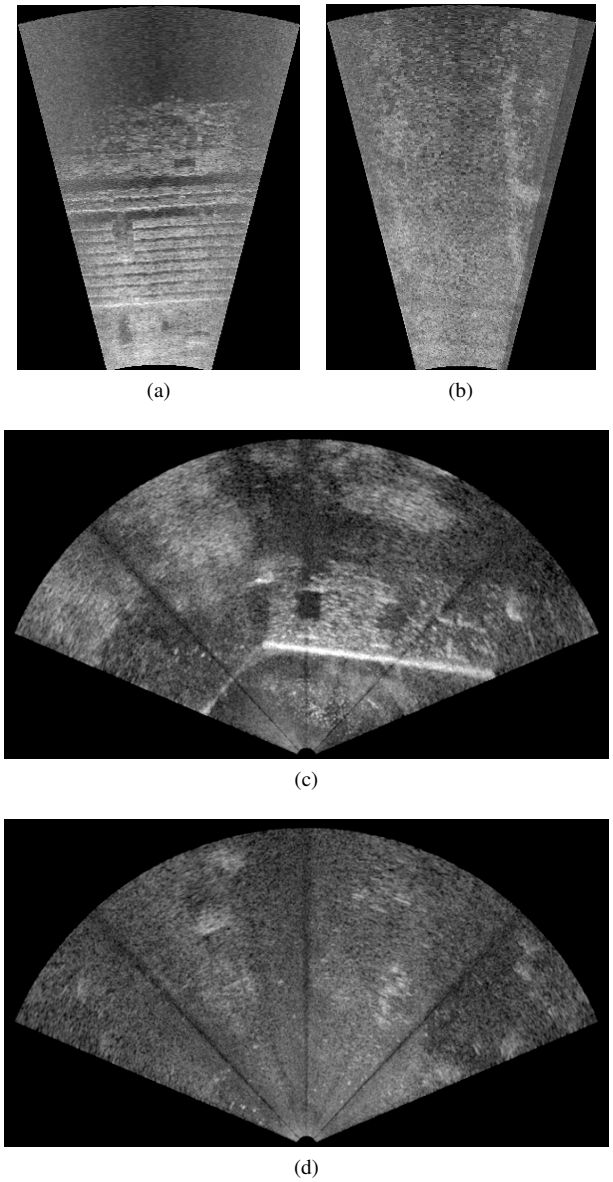


Fig. 2: Image examples of the experiment datasets. (a) DIDSON-ShipHull. (b) DIDSON-Natural environment. (c) BlueView-Dock. (d) BlueView-Natural environment.

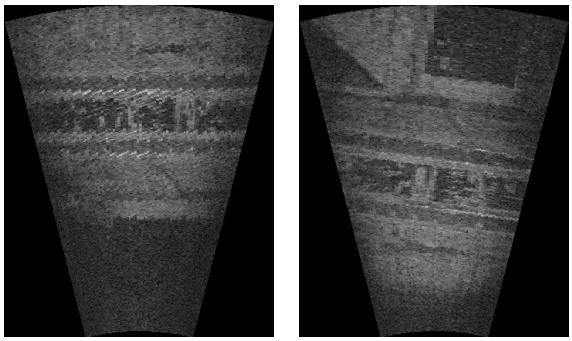


Fig. 3: Image example of a severe viewpoint change.

tion is the percentage of registrations that are successfully performed. We will count registrations as unsuccessful if the estimated rotation and/or translation differ from the ground-truth by a certain threshold, or if there were insufficient matches to estimate the underlying transformation. Although the translation errors can be correlated to rotation errors a threshold has been set separately. Those estimates that differ more than 3 degrees for the rotation and more than 5 pixels from the ground-truth translational displacements are labeled as unsuccessful registrations. These thresholds are fairly restrictive taking into account the resolution of the sonar images. In previous work [6] we have used a goodness measure tailored to the phase-correlation method to discern successful registrations, however, the thresholding criterion will be used here for comparison purposes. Additionally, the accuracy of the successful registrations has been analyzed by computing the mean squared error and the standard deviation of the rotation and translations with respect to the ground-truth estimates.

Finally, another parameter that will be taken into account in the evaluation of the results is the repeatability of the feature-detectors. Repeatability is an important characteristic of a feature detector, which indicates the ability to detect the same feature if the scene is revisited from a different viewpoint or there are illumination changes. In this case, we have computed the repeatability rate following the measure described in [22]. By knowing the homography that relates one image to the other, the detected features of the common image area are evaluated in a small neighborhood to determine whether they are repeated features or not. In order to prevent biased measures when the number of features is different in the two images, the measure is divided by the minimum number of detected features in the common area. In our tests we have set this neighborhood to 3 pixels.

V. RESULTS AND DISCUSSION

Figure 4 shows the percentage of successful registrations detected by each registration method in the first and second experiments. First, we have to highlight the performance of the phase-correlation approach, which is able to successfully register between 70% to 90% of the frames depending on the environment and the sonar's resolution and field of view. None of the other registration approaches are able to correctly register over 50% of the frames. The Harris corner detector performs reasonably well, particularly on the ship-hull inspection dataset where it clearly targets all the sharp illumination changes present in the pipes and other hull objects. The Hessian detector and MSER generally perform poorly, being just able to register successfully under our established thresholds around 15% and 10% of the sequences, showing a slightly better results in the case of natural environments. Although the difference is not enough to draw any solid conclusion it is indeed true that the dataset of natural environment images contain a higher number of blob-like features (rocks, small mountains, etc).

It should be noted that the Harris and Hessian detectors have been adjusted to detect a similar number of features,

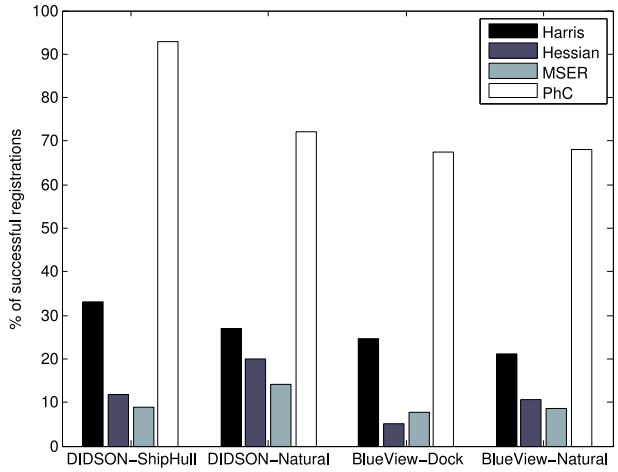


Fig. 4: Comparison of correctly registered FLS frames by different techniques under different environments and different sonar devices.

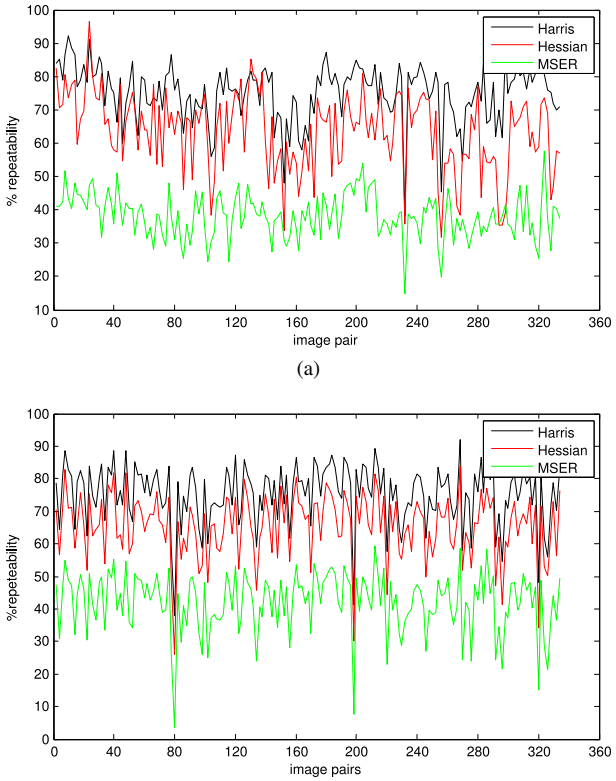


Fig. 5: Repeatability rates of the different feature detectors on the experiment (a)DIDSON-ShipHull. (b)DIDSON-Natural.

so that we do not introduce any bias in the comparisons. An average of 200 features are detected in the DIDSON datasets. Due to the larger field of view, more features are detected in BlueView datasets, on the order of 400-500. MSER detects fewer features in all cases. Therefore, even though the datasets imaging the seafloor lack the presence of strong and distinctive shapes, at pixel-level there are many features that can be still identified. However, specially in the case of BlueView datasets, a large number of detected features are located around small reflections and speckle noise. Since these speckle are spread all over the image, they can be matched with similar ones resulting in a significant number of matches that are coherent but erroneous, thus yielding incorrect registrations. This effect seems is less noticeable in DIDSON datasets, possibly due to its higher beamwidth resolution. As shown in figure 5, the repeatability rates of the detector are in accordance with the general registration results. Harris has the highest repeatability, of around 70%, followed by the Hessian and MSER. Repeatability rates for BlueView datasets follow the same trend.

Table I summarizes the mean squared error and standard deviation of the registrations that have been identified as correct according to our criteria. Since the correct registrations for Hessian and MSER detectors is very small, we analyze the results for the Harris and phase correlation methods. Once the failed registrations have been filtered out, we can see that errors are similar, although there are slightly larger rotation errors with phase correlation, probably due to the inaccuracies introduced by the polar transform of the Fourier Transform magnitude. The errors in the feature-rich BlueView dataset are slightly higher than its DIDSON counterpart, while in the natural environment are slightly smaller. This shows that, on one hand, a higher resolution definitely increases the accuracy of the registration while on the other hand, a larger field of view can compensate in those environments where there are few features. This is due to the fact that a large field of view allows the detection of a higher number and better spread features (for the feature-based techniques) or the inclusion of more image content to help disambiguating the correlation (for the Fourier-based registration), yielding a more accurate transformation estimate.

In the third experiment, we have analyzed the dataset containing 10 pairs of images with viewpoint changes. Neither Hessian nor MSER have been able to correctly align any of the 10 pairs. Harris corner detector has been able to align one, and the phase correlation algorithm has correctly recovered 4 transformations. Thus, all methods are sensitive to viewpoint changes. This vulnerability is not a problem of geometric complexity (the new viewpoint is still a combination of rotation and translation) but the difficulty resides in the photometric changes that are involved. If we compute the repeatability rates for these pairs (Fig. 6), it can be seen that are certainly lower, thus certifying the difficulty that the detectors have in recognizing the same feature under a viewpoint change.

Regarding the last experiment where the different methods are tested under several sonar artifacts, we should distinguish the particular kind of artifact to understand the results. In those pairs where there is a reflection from the water surface, a series of strong acoustic returns appear. When the affected area occupies a large portion of the image with respect to the actual imaged content (as is the case in the selected pairs), the detectors find multiple similar features

TABLE I: Estimated error mean and standard deviations for Harris and Phase correlation methods under the different datasets. Translations errors (t_x,t_y) are expressed in pixels and rotation errors (rot) in degrees.

		HARRIS	PHC
DIDSON-ShipHull	t_x	2.62 ± 0.7	2.91 ± 0.5
	t_y	1.92 ± 0.3	1.93 ± 0.4
	rot	1.33 ± 0.5	1.62 ± 0.4
DIDSON-Natural	t_x	4.0 ± 0.3	3.7 ± 0.5
	t_y	4.47 ± 0.5	4.29 ± 0.5
	rot	3.21 ± 0.5	3.72 ± 0.5
BlueView-Dock	t_x	3.44 ± 0.4	3.67 ± 0.4
	t_y	3.23 ± 0.5	3.91 ± 0.5
	rot	2.01 ± 0.6	2.43 ± 0.7
BlueView-Natural	t_x	3.57 ± 0.4	3.65 ± 0.4
	t_y	3.36 ± 0.5	3.23 ± 0.4
	rot	3.11 ± 0.3	3.56 ± 0.5

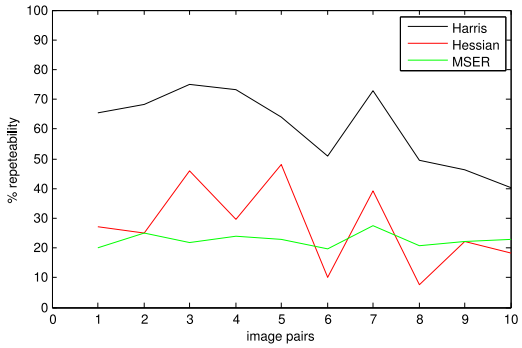
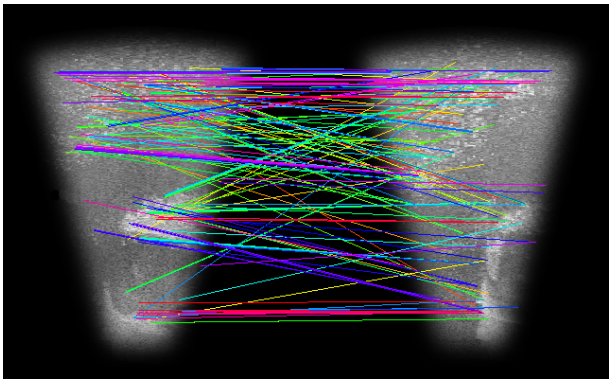
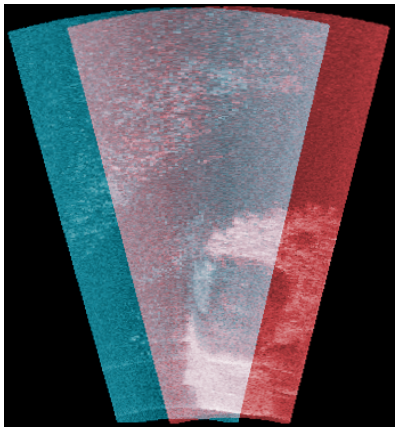


Fig. 6: Repeatability rates of the different feature detectors with image pairs from loop closing situations.



(a)



(b)

Fig. 7: Comparison in presence of water surface reflections (a) The feature detectors find multiple features on the reflections area (top left of the image), yielding to mismatches. (b) Phase-correlation achieves a correct registration.

in those areas, thus introducing mismatches that result in a failed alignment. With the phase correlation technique, the content of those reflections introduces noise to the phase correlation matrix, however since the actual image content is more coherent than the two varying reflections on the images, the correlation peak can still be determined and thus the images are correctly matched (Fig. 7). The cross-talk effects between beams, leading to the apparition of replicas of insonified structures in the images, does not seem to affect the registration techniques. Normally, the original structure is more strongly represented (in terms of image intensity) than the replicas, and the latter do not move coherently with the original structure from one pair to the other. Hence, the matches detected on the duplicates are removed correctly in the outlier rejection step. In the tested image pairs, Harris and phase correlation achieved correct registrations on all pairs while Hessian and MSER performed slightly worse (6 and 7 correct registrations respectively) although this cannot be linked to particular mismatches introduced by the cross-talk effects.

To summarize, the obtained results discourage the use of feature-based registration approaches at pixel-level since as it has been proved on real data, they are unable to deal with the low-resolution, noise, illumination changes and other artifacts present in the FLS imagery. On the other hand, phase-correlation performes well under all the tested circumstances even being able to register some loop-closing pairs, which is a key requirement to constrain the accumulated errors inside a global alignment framework. Concerning more practical aspects, Fourier-based registrations offer high computational efficiency due to the use of the Fast Fourier Transform (FFT) and its execution is determined by a constant time, linear with the size of the image, rather than the feature-based techniques whose processing time depends on the number of features detected at each frame.

VI. CONCLUSIONS

The development of 2D-FLS offer a brand new range of possibilities for both remotely-operated and autonomous underwater vehicles to operate in low visibility environments. To this end, it is required to find a suitable registration technique that can handle the difficult challenges of FLS images. We have presented a study of the most representative registration techniques according to a simplified sonar geometry model. Results show that the phase-correlation technique outperforms feature-based methods which have problems of repeatability and discrimination due to the low resolution and illumination changes. Future work will concentrate to further

adapt Fourier-based registration methods to the particularities of the FLS images, which given the potential shown in the present study are likely to become the standard of choice for the registration of this type of acoustic imagery.

REFERENCES

[1] S. Negahdaripour, H. Assalih, 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.

[2] 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.

[3] E. W. Yong, "Investigation of mosaicing techniques for forward looking sonar," Master's thesis, Heriot-Watt University, 2011.

[4] M. R. Walter, "Sparse bayesian information filters for localization and mapping," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, 2008.

[5] H. Johannsson, 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.

[6] N. Hurtos, X. Cufi, Y. Petillot, and J. Salvi, "Fourier-based registrations for two-dimensional forward-looking sonar imagery." IEEE, 2012, pp. 5298–5305, international Conference on Intelligent Robots and Systems.

[7] "Sound Metrics Corp." [Online]. Available: www.soundmetrics.com

[8] "Blueview Technologies Inc." [Online]. Available: www.blueview.com

[9] S. Negahdaripour, P. Firoozfam, and P. Sabzmeydani, "On processing and registration of forward-scan acoustic video imagery," in *Proc. 2nd Canadian Conf. Computer and Robot Vision*, 2005, pp. 452–459.

[10] C. Harris and M. Stephens, "A combined corner and edge detector," in *Alvey vision conference*, vol. 15. Manchester, UK, 1988, p. 50.

[11] 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.

[12] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[13] P. Biber and W. Straßer, "The normal distributions transform: A new approach to laser scan matching," in *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, vol. 3. Ieee, 2003, pp. 2743–2748.

[14] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: a survey," *Foundations and Trends® in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2008.

[15] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[16] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," *Computer Vision–ECCV 2006*, pp. 404–417, 2006.

[17] P. R. Beaudet, "Rotationally invariant image operators," in *Proceedings of the 4th International Joint Conference on Pattern Recognition*, Kyoto, Japan, Nov. 1978, pp. 579–583.

[18] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761–767, 2004.

[19] 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.

[20] 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.

[21] 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.

[22] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *International Journal of computer vision*, vol. 37, no. 2, pp. 151–172, 2000.