

# FORWARD-LOOKING SONAR MOSAICING FOR UNDERWATER ENVIRONMENTS

**Natàlia HURTÓS VILARNAU**

Dipòsit legal: Gi. 108-2015

<http://hdl.handle.net/10803/285086>

**ADVERTIMENT.** L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

**ADVERTENCIA.** El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

**WARNING.** Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.



Universitat de Girona

*Doctoral Thesis*

# **Forward-Looking Sonar Mosaicing for Underwater Environments**

**Natàlia HURTÓS VILARNAU**

2014





Universitat de Girona

*Doctoral Thesis*

# **Forward-Looking Sonar Mosaicing for Underwater Environments**

**Natàlia HURTÓS VILARNAU**

2014

*Doctoral Program in Technology*

*Supervised by:*

Joaquim Salvi, Xavier Cufí and Yvan Petillot

Thesis submitted to University of Girona in fulfillment of the requirements for the  
degree of

DOCTOR OF PHILOSOPHY



## CERTIFICATE OF THESIS DIRECTION

Dr. Joaquim Salvi and Dr. Xavier Cufí from Universitat de Girona, and Dr. Yvan Petillot from Heriot-Watt University,

DECLARE:

That the work entitled *Forward-Looking Sonar Mosaicing for Underwater Environments* presented by Natàlia Hurtós Vilarnau to obtain the degree in Doctor of Philosophy has been developed under our supervision and fulfills the requirements to obtain the International Mention.

Therefore, in order to certify the aforesaid statement, we sign this document.

*Girona, September 2014*

---

Dr. Joaquim Salvi

---

Dr. Xavier Cufí

---

Dr. Yvan Petillot



*Al meu pare,  
cause the time of flight is not always  
proportional to how far one reaches.*





# ACKNOWLEDGMENTS

---

Nothing better than the topic of this thesis to know that when you don't see things clearly you should listen. The PhD journey is truly a case where many (and very different!) things become blur along the way, but I have been fortunate enough to be surrounded of incredible people to hear out. Along these lines I would like to express my gratitude to all of them, family, friends and work mates, that helped me in countless ways.

M'agradaria començar agraïnt el suport dels meus pares que sempre m'han animat a seguir estudiant i m'han educat amb uns valors sense els quals hauria sigut difícil arribar fins aquí. Desafortunadament, el meu pare no podrà llegir aquestes línies però el seu esforç, constància i valentia, han estat i seran sempre un exemple a seguir per mi. A la meva mare li dec més que un agraïment, perquè mentre jo m'he dedicat a allò que ella resumeix com "*enganxar fotos de sota l'aiga*", s'ha ocupat (i preocupat) d'un munt de coses, a banda de preparar tot sovint fantàstics *tuppers*. També he de donar les gràcies a la meva germana, els meus nebots i la resta de la família que s'han encarregat regularment d'amenitzar les meves estones lliures entre dinars, sortides i celebracions i més d'una vegada fer-me de taxi cap a l'aeroport.

The biggest of my thanks are for Aggelos, for who I could have words in every category of these acknowledgements, but he deserves a unique one, as unique as his support, advice and understanding throughout these years. He turned out to be not only a good travel companion for conferences but the best one to share every-day's struggles and laughs. And that is, no matter the contributions reported in next chapters, the greatest of my findings in this PhD.

I encara abans d'entrar en l'apartat de feina, vull agrair a tots els amics/gues que durant aquests anys han estat sempre disposats a fer "activitats extraescolars" i xefles variades, tant necessàries per recordar que el que de debó importa està per sobre l'aigua i no surt als *papers*.

Ara sí, he de donar les gràcies als meus supervisors. Primer a en Quim i a en Xevi, que un bon dia em van oferir la possibilitat de començar aquest doctorat i que m'han donat la llibertat suficient i l'orientació necessària per tirar-ho endavant. L'un tant pràctic i l'altre tant entusiasta, espero que se m'hagi encomanat quelcom al llarg d'aquests anys. Finalment, a l'Yvan, to who I owe a big thank you for introducing me the topic of this thesis and encouraging me since then.

---

Les gràcies més merescudes per tots els meus companys del CIRS. Els que fa anys que aguanten la paradeta, els més nous i també els que han passat i ara es troben en algun racó o altre de món. Des de les banyades estiuenques, les profundes converses político-futbolístiques de l'hora de dinar, els partits de pàdel o les cerveses a barraques, podria afirmar que tinc la sort que els meus companys de laboratori siguin també amics, si no fos, es clar, perquè amb els amics no acostumo a entrar robots a les habitacions dels hotels, enviar-ne d'altres a Kazakhstan o treballar en una cadena de producció per muntar 3 AUVs en una setmana. Voldria agrair molt especialment a en Pere, en Marc, en Narcís, en David, en Nuno i a en Lluís la seva ajuda en tantes i tantes coses (i de naturalesa tant diversa) que si les hagués d'escriure aquí m'ocuparia més que la resta d'aquest document. Ho resumiré en un gràcies per la vostra predisposició, paciència, i per, sovint, creure més en mi que jo mateixa. També un agraïment especial als més joves: l'Arnau, en Guillem, l'Albert, en Juan David, en Jep, en Carles i l'Eduard, que fan que al laboratori hi hagi un *remix* de neurones, caràcters i aficions tan singular com divertit. I finalment a aquells amb qui he coincidit al CIRS i m'han ajudat amb savis consells abans d'emigrar: l'Andrés, l'Emili, l'Enric, l'Enrique i en Sharad. Tampoc em voldria oblidar de la resta de l'equip amb qui vaig compartir èxit al SAUC-E 2010: en Ricard, en Miki, en Carlos, en Simone, en Sebas, en Chee-Sing, en Xevi, en Simó i en Pere. Evidentment, també vull extendre aquest agraïment a la gent del P4 (Rafa, Ricard, Quintana, Aulinas, Tudor,...) i a tot el grup VICOROB en general. I com no, a les *secres* més guapes i trempades de tot el món universitari: la Joseta, la Mireia i les Annes. Ah, i en Jordi! Així doncs, a tots, les meves gràcies més sinceres. I com marca la tradició, no patiu, també tindreu les meves gràcies en format comestible.

I also owe an acknowledgement to the colleagues of the Ocean Systems Lab in Edinburgh, with whom I shared the biggest Scottish snowfall in years and many bags of shortbread cookies (Francesco deserves especial thanks for that) while I was doing my first steps in this PhD.

I am also grateful to the people that have shared with us their sonar data, which allowed me to keep working, especially when no forward-looking sonar was available at CIRS. In this regard, I should thank SoundMetrics Corp., Bluefin Robotics Corp., SeeByte Ltd., Vladimir Djapic, and all the people that has showed interest in our work and have sent us their files for evaluation. That really helped me to gain a great deal of experience and made me feel gratified seeing that the work I have been doing can also be useful outside of the lab.

I want to express also my gratitude to the anonymous reviewers that contributed with valuable comments to improve the work and the publications derived from this PhD. Likewise I would like to thank the reviewers of this thesis manuscript for their time and their comments, and Judy Fenwick for proofreading it faster and better than I could have ever imagined. Finally I would like to thank the agencies that partially funded this work: the Spanish Ministry of Science and Innovation under the FPI grant DPI2007-66796-C03-02 (AQUAVISION), the national projects CTM2011-29691-C02-02 RAIMON and DPI2011-27977-C03-02 COMAROB, and the EU projects FP7-ICT-2011-7-288273 PANDORA and FP7-ICT-2009-248497 TRIDENT.

# LIST OF PUBLICATIONS

---

## Publications derived from this thesis

The work developed in this thesis led to the following publications:

- **N. Hurtós**, D. Ribas, X. Cufi, Y. Petillot, and J. Salvi, “Fourier-based registration for robust forward-looking sonar mosaicing in low-visibility underwater environments,” *Journal of Field Robotics*, 2014.
- **N. Hurtós**, S. Nagappa, N. Palomeras, and J. Salvi, “Real-time mosaicing with two-dimensional forward-looking sonar,” in *International Conference in Robotics and Automation (ICRA)*, Hong-Kong, 2014.
- **N. Hurtós**, N. Palomeras, M. Carreras, A. Carrera, C. Bechlioulis, G. Karras, S. Heshmati-alamdari, and K. Kyriakopoulos, “Sonar-based chain following using an autonomous underwater vehicle,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems, (IROS)*, Chicago, Illinois, 2014.
- **N. Hurtós**, X. Cufi, Y. Petillot, and J. Salvi, “Fourier-based registrations for two-dimensional forward-looking sonar image mosaicing,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems, (IROS)*. Vilamoura, Portugal: IEEE, 2012, pp. 5298–5305.
- **N. Hurtós**, “Forward-looking sonar mosaicing for underwater vehicles in low-visibility conditions,” in *RSS Women in Robotics Workshop.*, Berkeley, California, 2014.
- **N. Hurtós**, X. Cufi, and J. Salvi, “Rotation estimation for two-dimensional forward-looking sonar mosaicing,” in *ROBOT2013: First Iberian Robotics Conference. Advances in Intelligent Systems and Computing*, M. A. Armada, A. Sanfeliu, and M. Ferre, Eds. Madrid, Spain: Springer International Publishing, 2014, vol. 252, pp. 69–84.
- **N. Hurtós**, X. Cufi, and J. Salvi, “A novel blending technique for two dimensional forward looking sonar mosaicking,” in *IEEE/MTS OCEANS - San Diego*, San Diego, California, 2013.
- **N. Hurtós**, S. Nagappa, X. Cufi, Y. Petillot, and J. Salvi, “Evaluation of registration methods on two-dimensional forward-looking sonar imagery,” in *MTS/IEEE OCEANS - Bergen*, Bergen,

Norway, 2013.

- **N. Hurtós**, N. Palomeras, and J. Salvi, “Automatic detection of underwater chain links using a forward-looking sonar,” in *MTS/IEEE OCEANS - Bergen*, Bergen, Norway, 2013.

## Other publications

Parallel work at the time of this thesis led to the following publications:

- A. Carrera, G. Karras, C. Bechlioulis, N. Palomeras, **N. Hurtós**, K. Kyriakopoulos, P. Kormushev, and M. Carreras, “Improving a learning by demonstration framework for intervention auvs by means of an UVMS controller,” in *ICRA Workshop on Persistent Autonomy for Marine Robotics*, Hong Kong, 2014.
- A. Carrera, N. Palomeras, **N. Hurtós**, P. Kormushev, and M. Carreras, “Learning by demonstration applied to underwater intervention,” in *17th International Conference of the Catalan Association of Artificial Intelligence*, Barcelona, 2014.
- P. Ridao, D. Ribas, N. Palomeras, M. Carreras, A. Mallios, **N. Hurtós**, N. Gracias, L. Magí, R. Garcia, R. Campos *et al.*, “Operational validation of Girona500 AUV,” in *5th International Workshop on Marine Technology (Martech)*. Girona, Spain: SARTI, 2013.
- P. Nomikou, J. Escartin, P. Ridao, D. Sakellariou, R. Camilli, V. Ballu, M. Moreira, C. Mevel, A. Mallios, C. Deplus, M. Andreani, O. Pot R. Garcia, L. Rouzie, T. Gabsi, R. Campos, N. Gracias, **N. Hurtós**, L. Magí, N. Palomeras, and D. Ribas, “Preliminary submarine monitoring of santorini caldera: Hydrothermal activity, and seafloor deformation.” in *Volcanism of the Southern Aegean in the frame of the broader Mediterranean area (VOLSAM)*, 2012.
- **N. Hurtós**, X. Cufi, and J. Salvi, “Calibration of optical camera coupled to acoustic multibeam for underwater 3D scene reconstruction,” in *OCEANS 2010 IEEE - Sydney*, Sydney, Australia, 2010, pp. 1–7.
- **N. Hurtós**, X. Cufi, and J. Salvi, “Integration of optical and acoustic sensors for 3D underwater scene reconstruction.” in *3rd International Workshop on Marine Technology (Martech)*, SARTI, Ed., no. 8, Vilanova i la Geltru, Spain, 2009.

# ACRONYMS

---

**2D** two-dimensional

**3D** three-dimensional

**AHRS** Attitude and Heading Reference System

**ASV** Autonomous Surface Vehicle

**AUV** Autonomous Underwater Vehicle

**BA** Bundle Adjustment

**CASC** Catalan Centre for Underwater Archaeology, *Centre d'Arqueologia Subaquàtica de Catalunya*

**CIRS** Underwater Robotics Research Center, *Centre d'Investigació en Robòtica Submarina*

**CLAHE** Contrast Limited Adaptive Histogram Equalization

**CMRE** Centre for Maritime Research and Experimentation

**DOF** Degree of Freedom

**DFT** Discrete Fourier Transform

**DVL** Doppler Velocity Log

**ESEIF** Exactly Sparse Extended Information Filter

**FFT** Fast Fourier Transform

**FLS** Forward-Looking Sonar

**FOV** Field of View

**FPSO** Floating Production Storage and Offshore

**G2O** General Framework for Graph Optimization

**GPS** Global Positioning System

**HAUV** Hovering Autonomous Underwater Vehicle

**IFFT** Inverse Fast Fourier Transform

---

**MRI** Magnetic Resonance Imaging  
**NDT** Normal Distribution Transform  
**OSL** Ocean Systems Lab  
**PCL** Point Cloud Library  
**PSR** Peak-to-Sidelobe Ratio  
**RANSAC** RANdom SAmple Consensus  
**ROS** Robot Operative System  
**ROV** Remotely Operated Vehicle  
**RTK** Real-Time Kinematic  
**SAUC-E** Student Autonomous Underwater Challenge - Europe  
**SIFT** Scale-Invariant Feature Transform  
**SURF** Speeded-Up Robust Features  
**SLAM** Simultaneous Localization And Mapping  
**SNR** Signal-to-Noise Ratio  
**TVG** Time Varying Gain  
**UVL** Underwater Vision Lab

# LIST OF FIGURES

---

1.1	Example of procedures requiring the underwater acquisition of visual data. . . .	3
1.2	Different examples of optical images affected by backscatter and/or turbid water conditions. . . . .	4
1.3	Different underwater sonar types and its data. . . . .	6
1.4	Typical mosaicing pipeline. . . . .	7
1.5	Example of Forward-Looking Sonar (FLS) mosaics reported in previous works. .	8
1.6	Chapter organization outline. . . . .	11
2.1	FLS operation. . . . .	14
2.2	Sonar projection geometry. . . . .	17
2.3	Projection error due to the orthographic approximation. . . . .	17
2.4	Effect of rotations and translations on FLS images . . . . .	18
2.5	Elevation angle estimation from leading and trailing image edges. . . . .	21
2.6	Feature shadow detection for elevation estimation. . . . .	22
2.7	Inhomogeneous resolution in FLS images . . . . .	23
2.8	Inhomogeneous insonification in FLS images . . . . .	24
2.9	Other artifacts present in FLS images . . . . .	24
3.1	Feature-based registration pipeline. . . . .	27
3.2	Feature-based registration at region level. . . . .	29
3.3	Basic phase correlation workflow. . . . .	31
3.4	Example of the Balci and Foroosh method. . . . .	32
3.5	Phase correlation workflow with masking. . . . .	33
3.6	Image masking procedure. . . . .	33
3.7	Masking of edge effects. . . . .	34
3.8	Examples of obtained peak surfaces when applying phase correlation to overlapping FLS images. . . . .	35
3.9	Diagrams of prefiltering and postfiltering operations. . . . .	36



---

3.10	Effect of the averaging filter applied on the correlation surface for different kernel sizes. . . . .	37
3.11	Butterworth filter. . . . .	38
3.12	two-dimensional (2D) Delta and its Fast Fourier Transform (FFT). . . . .	39
3.13	Adaptive determination of the filter cutoff frequency. . . . .	40
3.14	Examples of automatically determined $f_c$ values and their Peak-to-Sidelobe Ratio (PSR) values compared to a range of possible filter cutoff frequencies. . . . .	41
3.15	Contour plot of the angular difference function. . . . .	44
3.16	Distortions on the estimation of the rotation angle as a shift in the polar images. . . . .	46
3.17	Overall registration pipeline. . . . .	47
3.18	Example of dataset frames used in the quantitative comparisons . . . . .	51
3.19	Example frames of a featureless dataset. . . . .	54
4.1	Depiction of the pose-based graph structure . . . . .	60
4.2	Interaction between graph front-end and back-end. . . . .	62
4.3	Representation of the proposed heuristic to compute the uncertainty of the registration from the phase correlation matrix values. . . . .	63
4.4	Schematics illustrating the performance of the proposed heuristic for estimating the registration uncertainty versus Pfingsthorn <i>et al.</i> heuristic. . . . .	65
4.5	Uncertainty of an unsuccessful registration not containing the true registration point. . . . .	65
4.6	Flow diagram of the global alignment for the offline approach. . . . .	69
4.7	Flow diagram of the global alignment for the online approach. . . . .	71
5.1	Example of the denoising effect obtained by mosaicing . . . . .	76
5.2	Correction of non-inhomogeneous insonification pattern. . . . .	78
5.3	Image equalisation and contrast enhancement. . . . .	78
5.4	Example of dataset frames affected by blind areas and varying illumination depending on the imaging viewpoint. . . . .	80
5.5	Mosaic blending corrections at global level. . . . .	82
6.1	Tank experiment setup . . . . .	84
6.2	Trajectories and graph of the ARIS Tank experiment . . . . .	85
6.3	Mosaic of the tank test. . . . .	86
6.4	Detail comparison of a small area in the tank mosaic . . . . .	86
6.5	Hovering Autonomous Underwater Vehicle (HAUV) vehicle with DIDSON FLS used in the ship hull experiment. . . . .	87
6.6	Links established by registration constraints in the ship hull dataset . . . . .	88
6.7	Trajectories of the ship hull dataset . . . . .	89
6.8	Ship hull mosaic rendered over different trajectories . . . . .	90

6.9	Detail comparison of a small area in the ship hull mosaic. . . . .	91
6.10	Qualitative comparison of the mosaics elaborated from two different datasets gathered on the same ship hull. . . . .	91
6.11	Autonomous Surface Vehicle (ASV) equipped with the Blueview P900-130 FLS used in the marina marciana experiment. . . . .	92
6.12	Absolute mean errors of the registration estimates for consecutive frames in the Marciana Marina dataset . . . . .	94
6.13	Error histograms for consecutive frame registrations in the Marciana Marina dataset. . . . .	95
6.14	Loop closure links detected in the Marciana Marina experiment. . . . .	95
6.15	Trajectories of the Marciana Marina experiment: GPS trajectory (in green) and estimated trajectory after the global alignment (in red). . . . .	96
6.16	Mosaic and orthophotomap of the Marciana Marina experiment . . . . .	97
6.17	<i>Cap de Vol</i> experiment . . . . .	98
6.18	<i>Cap de Vol</i> experiment trajectories . . . . .	100
6.19	<i>Cap de Vol</i> experiment graph after running the online mosaicing. . . . .	100
6.20	<i>Cap de Vol</i> shipwreck mosaic. . . . .	101
6.21	Portion of the <i>Cap de Vol</i> mosaic with the overlaid planimetry. . . . .	102
A.1	Different imaging configurations. . . . .	112
A.2	Imaging with large altitude and large sonar tilt angle. . . . .	113
A.3	Imaging with large altitude and large ranges. . . . .	113
A.4	Imaging with large altitude and insufficiently large ranges. . . . .	114
A.5	Imaging with low altitude. . . . .	114
A.6	Imaging with large tilt . . . . .	114
A.7	Imaging with small tilt. . . . .	115
A.8	Imaging with loose ranges. . . . .	115
A.9	Imaging with adjusted ranges. . . . .	115
A.10	Grid surveys maintaining FLS orientation. . . . .	117
B.1	Diagram illustrating the different steps of the chain following framework. . . . .	122
B.2	Diagram of the link detection steps. . . . .	124
B.3	Phases of the waypoint planner computations . . . . .	126
B.4	Chain experiment setup at the University of Girona's water tank . . . . .	128
B.5	Results of the chain following experiment . . . . .	130
B.6	Link waypoints overlaid on mosaiced FLS data . . . . .	130
B.7	Example of chain mosaic. . . . .	132



# LIST OF TABLES

---

2.1	Specifications of main FLS devices present in the market during the elaboration of this thesis. . . . .	16
3.1	Comparison experiments between rotation estimation methods when registering consecutive frames. . . . .	49
3.2	Comparison experiments between rotation estimation methods when registering distant frames. . . . .	49
3.3	Results of the feature-based registration at pixel level when registering consecutive frames. . . . .	50
3.4	Results of the feature-based registration at pixel level when registering distant frames. . . . .	51
3.5	Comparison experiments between the Fourier-based and region-based registration methods when registering consecutive frames. . . . .	53
3.6	Comparison experiments between the Fourier-based and region-based registration methods when registering distant frames. . . . .	53
4.1	Evaluation of uncertainty measures . . . . .	64
4.2	Evaluation of uncertainty measures only for successful registrations . . . . .	66
6.1	Summary table for the Tank Test Dataset. . . . .	84
6.2	Summary table for the ship hull dataset. . . . .	88
6.3	Summary table for the Marina Marciana dataset. . . . .	93
6.4	Summary table for the <i>Cap de Vol</i> dataset. . . . .	98
B.1	Table showing the number and accuracy (compared to manually labelled link centers) of the link detections along the chain following experiment. . . . .	129



# CONTENTS

---

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	2
1.2	Objectives . . . . .	8
1.3	Context . . . . .	9
1.4	Document Roadmap . . . . .	10
<b>2</b>	<b>Background on Forward-Looking Sonar Imagery</b>	<b>13</b>
2.1	FLS operation . . . . .	14
2.2	Imaging geometry model . . . . .	15
2.2.1	Orthographic projection approximation . . . . .	15
2.2.2	Non-approximated projection . . . . .	19
2.2.3	Selecting a FLS geometry model for 2D mosaicing . . . . .	20
2.3	Challenges in FLS imagery . . . . .	21
<b>3</b>	<b>Registration of Forward-Looking Sonar Images</b>	<b>25</b>
3.1	Related work on FLS Image Registration . . . . .	26
3.1.1	Feature-based registration at pixel level . . . . .	26
3.1.2	Feature-based registration at region level . . . . .	28
3.1.3	Taking it a step further: area-based registration . . . . .	28
3.2	Fourier-based registration for FLS . . . . .	29
3.2.1	Phase correlation for the estimation of translational shifts . . . . .	30
3.2.2	Rotation estimation . . . . .	41
3.2.3	FLS Registration Outline . . . . .	47
3.3	Quantitative Comparisons . . . . .	47
3.3.1	Comparison of the simplified rotation estimation method versus the Fourier-Mellin Transform . . . . .	48
3.3.2	Comparison against feature-based registration . . . . .	49
3.3.3	A note on computational cost . . . . .	54

---

<b>4</b>	<b>Global Alignment</b>	<b>57</b>
4.1	Related work . . . . .	58
4.2	Pose-based Graph Definition . . . . .	59
4.3	Estimation of the Constraints Uncertainty . . . . .	62
4.3.1	Registration Constraints . . . . .	62
4.3.2	Navigation Constraints . . . . .	66
4.4	Graph Construction . . . . .	67
4.4.1	Offline approach . . . . .	67
4.4.2	Real-time approach . . . . .	69
<b>5</b>	<b>Mosaic Rendering</b>	<b>73</b>
5.1	Blending of acoustic images . . . . .	74
5.2	Intensity averaging blending . . . . .	75
5.3	Improving the blending of FLS mosaics . . . . .	76
5.3.1	Individual image pre-processing . . . . .	77
5.3.2	Global mosaic blending . . . . .	79
5.4	Resolution enhancement . . . . .	81
<b>6</b>	<b>Experiments and Results</b>	<b>83</b>
6.1	Tank test . . . . .	84
6.1.1	Experiment description . . . . .	84
6.1.2	Results . . . . .	84
6.2	Ship Hull Inspection . . . . .	87
6.2.1	Experiment description . . . . .	87
6.2.2	Results . . . . .	88
6.3	Marina Marciana Harbor Mapping . . . . .	92
6.3.1	Experiment description . . . . .	92
6.3.2	Results . . . . .	93
6.4	<i>Cap de Vol</i> Archaeological Site . . . . .	96
6.4.1	Experiment description . . . . .	96
6.4.2	Results . . . . .	98
<b>7</b>	<b>Conclusions</b>	<b>103</b>
7.1	Summary of completed work . . . . .	104
7.2	Review of main contributions . . . . .	106
7.3	Framework failures . . . . .	106
7.4	Future work . . . . .	108
<b>A</b>	<b>Guidelines for FLS image acquisition</b>	<b>111</b>
A.1	Appropriateness of the scenario . . . . .	112

---

A.2	Guidelines for imaging a horizontal plane . . . . .	112
A.2.1	Sonar imaging setup . . . . .	113
A.2.2	Other considerations . . . . .	116
<b>B</b>	<b>Chain detection and following on FLS imagery</b>	<b>119</b>
B.1	Motivation . . . . .	120
B.2	System Overview . . . . .	121
B.3	Detection of Chain Links on FLS imagery . . . . .	122
B.4	Way-Point Planner . . . . .	124
B.5	Control Scheme . . . . .	126
B.6	Experiments and Results . . . . .	127
	<b>Bibliography</b>	<b>146</b>





# ABSTRACT

---

Vehicle operations in underwater environments are frequently compromised by poor visibility conditions. The perception range of optical devices is heavily constrained in turbid waters, thus often complicating navigation and mapping tasks in environments such as harbors, bays, or rivers. A new generation of high-frequency forward-looking sonars that provide acoustic imagery at near-video frame rates have recently emerged as a promising alternative for working under these challenging conditions.

In this thesis, we propose an end-to-end mosaicing framework tailored to the characteristics of forward-looking sonar imagery in order to build consistent overviews of planar underwater areas regardless of water visibility. Our solution targets versatility: it enables the generation of acoustic mosaics that involve roto-translational motions and comprise different vehicle tracklines; it is suitable for a wide range of scenarios, from feature-rich areas to environments with scarcity of features; it can be applicable on data collected with minimally instrumented vehicles; and it allows both offline and real-time operation.

The first problem to address is the pairwise registration of sonar images which is a key step in the mosaicing pipeline. The characteristics of the forward-looking sonar data, such as low and inhomogeneous resolution, low signal-to-noise ratio and intensity variations due to viewpoint changes, become a challenge for traditional feature-based registration techniques. For that reason we propose a Fourier-based methodology that, by involving all image content into the registration, offers robustness to noise and the different artifacts associated with the acoustic image formation. The approach relies on the phase correlation principle to estimate the image shifts and it is further adapted to cope with the multiple noise sources that can influence the registration, by providing specific masking, frequency filtering and rotation estimation procedures. When quantitatively compared, the proposed registration method shows superior performance to state-of-the-art feature-based approaches, while offering at the same time the possibility to be implemented efficiently.

Next, we address the global alignment of the mosaic in order to enforce consistency between consecutive and non-consecutive image pairs. We lay out the problem as a graph optimization over the image poses, integrating spatial constraints from pairwise registrations as well as from

navigation data when available. We provide a front-end to determine the constraints that should be included in the graph according to an initial estimation of the trajectory and a selection of potential overlapping candidate pairs. The workflow followed to build the graph is provided for both offline mosaicing and the online approach, where constraints are added incrementally and under stringent restrictions to warrant real-time operation. In addition, we propose an uncertainty measure derived from the registration method to weigh appropriately the contribution of the registration constraints in the optimization.

Finally, we explore the blending of the acoustic images into a smooth and informative mosaic while improving the signal-to-noise ratio and resolution of the final composition with respect to the individual frames. Furthermore, we identify the different photometric irregularities that can arise from the sonar imaging configuration and provide a set of strategies to minimize their impact both at frame and mosaic level.

To validate the full proposed pipeline, an extensive experimental section is reported showing successful results in relevant field applications such as ship-hull inspection, harbor mapping and archaeological exploration.

# RESUM

---

Freqüentment, les operacions amb vehicles en entorns submarins estan condicionades a la visibilitat de l'aigua. El rang de percepció dels sensors òptics és molt limitat en aigües tènues i sovint fa complicades les tasques de navegació i mapeig en entorns com ara ports, badies o rius. Recentment han aparegut al mercat una nova generació de sonars de visió frontal, d'alta freqüència i amb la capacitat de generar imatges acústiques a un alt ritme de refresc, convertint-se així en una alternativa prometedora per a operar en aquestes difícils condicions.

En aquesta tesi, proposem un sistema complet per a la construcció de mapes subaquàtics adaptat a les característiques d'aquest tipus de sonars, per tal de construir mosaics d'imatges acústiques independentment de la visibilitat de l'aigua. La solució proposada té la versatilitat com a eix central: permet la generació de mosaics acústics que involucren moviments roto-translacional i múltiples transectes, és adequada per un ampli ventall d'escenaris, des d'àrees amb alta presència de punts d'interès fins a entorns mancats de característiques significants, és aplicable a dades adquirides des de vehicles equipats amb poca instrumentació i permet la generació de mosaics tant en mode post-processat com en temps real.

El primer problema que s'adreça és el registre d'imatges sonar, que és un pas clau en el sistema de construcció de mosaics. Les característiques de les imatges dels sonars de visió frontal, com ara la baixa resolució, la baixa relació senyal-soroll, o les variacions d'intensitat a causa dels canvis de punt de vista, causen moltes complicacions a les tècniques que típicament s'utilitzen per registrar imatges òptiques. Per aquesta raó proposem un mètode basat en el domini freqüencial, el qual té en compte tot el contingut de les imatges en el procés de registre aconseguint així més robustesa davant del soroll i dels diferents artefactes associats amb la formació de la imatge acústica. El mètode es basa en el principi de la correlació de fase per computar els desplaçaments de la imatge i s'ha adaptat per fer front a les múltiples fonts de soroll que poden influenciar el registre proporcionant procediments específics d'emascament, filtratge freqüencial i estimació de la rotació. Comparacions quantitatives demostren que el mètode proposat té un rendiment superior a altres tècniques de registre de l'estat de l'art basades en punts d'interès i ofereix, al mateix temps, la possibilitat de ser implementat eficientment.

A continuació s'ha adreçat l'alineament global del mosaic per tal d'imposar consistència entre

parelles d'imatges consecutives i no consecutives. S'ha plantejat el problema com la optimització d'un graf on els nodes són les posicions de les imatges i les restriccions espacials entre ells provenen del registre de les diferents parelles o de les dades de navegació, en cas de disposar d'aquestes. Detallem l'estratègia a seguir per determinar quines restriccions s'han d'incorporar al graf d'acord amb una estimació inicial de la trajectòria del vehicle i una selecció de les parelles d'imatges que poden tenir solapament. El procediment per a la construcció del graf es descriu tant per l'elaboració de mosaics *offline* com *online*. En aquest darrer cas, les restriccions s'afegeixen incrementalment i sota estrictes límits per garantir l'execució en temps real. A més a més, proposem una mesura d'incertesa derivada del mètode de registre per tal de donar el pes adequat a cada restricció dins de la optimització.

Finalment, s'explora la fusió de les imatges acústiques en un únic mosaic d'aparença nítida i informativa, aconseguint al mateix temps una relació senyal-soroll i una resolució millorades respecte les de les imatges individuals. També s'identifiquen les diferents irregularitats fotomètriques que poden sorgir degut a la configuració i col·locació del sonar en l'entorn i es proporcionen un seguit d'estratègies per minimitzar el seu impacte tant a nivell d'imatge com a nivell de mosaic.

Per tal d'avaluar la totalitat del marc de treball proposat, es presenta una extensiva secció experimental demostrant resultats satisfactoris en aplicacions rellevants com ara inspecció de cascos de vaixells, mapeig de ports o exploracions arqueològiques.

# RESUMEN

---

Frecuentemente, las operaciones con vehículos en entornos submarinos se ven condicionadas a la visibilidad del agua. El rango de percepción de los sensores ópticos es muy limitado en aguas turbias y a menudo complican las tareas de navegación y mapeo de entornos tales como puertos, bahías o ríos. Recientemente han aparecido en el mercado una nueva generación de dispositivos sonar de visión frontal, de alta frecuencia y con la capacidad de generar imágenes acústicas a una alta tasa de actualización, convirtiéndose así en una alternativa prometedora para operar en estas difíciles condiciones.

En esta tesis, proponemos un sistema completo para la construcción de mapas submarinos adaptado a la características de este tipo de sónars, con el objetivo de construir mosaicos de imágenes, independientemente de cual sea la visibilidad del agua. La solución que proponemos tiene la versatilidad como eje central: permite la generación de mosaicos acústicos que involucren movimientos roto-translacionales y múltiples transectos; es adecuada para un amplio rango de escenarios, desde áreas con alta presencia de puntos de interés hasta entornos faltos de características significantes; es aplicable en datos adquiridos desde vehículos equipados con poca instrumentación y permite la generación de mosaicos tanto en modo pos-procesado como en tiempo real.

El primer problema que se trata es el registro de imágenes de sonar, que constituye un paso clave en el sistema de construcción de mosaicos. Las características de las imágenes de un sonar de visión frontal, como la baja resolución, la baja relación señal-ruido o las variaciones de intensidad debido a los cambios de punto de vista, suponen muchas complicaciones a las técnicas que típicamente se usan en el registro de imágenes ópticas. Por esta razón, proponemos un método basado en el dominio frecuencial, el cual tiene en consideración todo el contenido de las imágenes en el proceso de registro consiguiendo así mejor robustez contra el ruido y los distintos artefactos asociados con la formación de la imagen acústica. El método se basa en el principio de la correlación de fase para computar los desplazamientos de la imagen y se ha adaptado para hacer frente a las múltiples fuentes de ruido que pueden influenciar el registro, proporcionando procedimientos específicos de enmascaramiento, filtraje frecuencial y estimación de la rotación. Comparaciones cuantitativas demuestran que el método propuesto tiene un rendimiento superior

a otras técnicas de registro del estado del arte basadas en puntos de interés y ofrece, al mismo tiempo, la posibilidad de ser implementado eficientemente.

A continuación se ha tratado el alineamiento global del mosaico con el fin de imponer consistencia entre parejas de imágenes consecutivas y no consecutivas. Se ha planteado el problema como la optimización de un grafo donde los nodos son las posiciones de las imágenes y las restricciones espaciales que se integran entre ellos provienen del registro de las distintas parejas o de los datos de navegación en el caso que se disponga de ellos. Detallamos la estrategia a seguir para determinar cuáles son las restricciones que se deben incorporar al grafo de acuerdo con una estimación inicial de la trayectoria del vehículo y una selección de las parejas de imágenes que pueden tener solapamiento. El procedimiento para la construcción del grafo se describe tanto para la elaboración de mosaicos *offline* como *online*. En este último caso, las restricciones se añaden incrementalmente y bajo estrictos límites para garantizar la ejecución en tiempo real. Además, proponemos una medida de incertidumbre derivada del método de registro con el fin de dar el peso adecuado a cada restricción dentro de la optimización.

Finalmente, se explora la fusión de las imágenes acústicas en un único mosaico de apariencia nítida e informativa, consiguiendo, al mismo tiempo, una relación señal-ruido y una resolución mejoradas respecto a las imágenes individuales. También se identifican las diferentes irregularidades fotométricas que pueden aparecer debido a la configuración y la colocación del sónar en el entorno y se proporcionan una serie de estrategias para minimizar su impacto tanto a nivel de imagen como a nivel de mosaico.

Para evaluar la totalidad del marco de trabajo propuesto, se presenta una extensiva sección experimental demostrando resultados satisfactorios en aplicaciones relevantes como inspección de casco de barcos, mapeo de puertos o exploraciones arqueológicas.

# 1

## INTRODUCTION

---

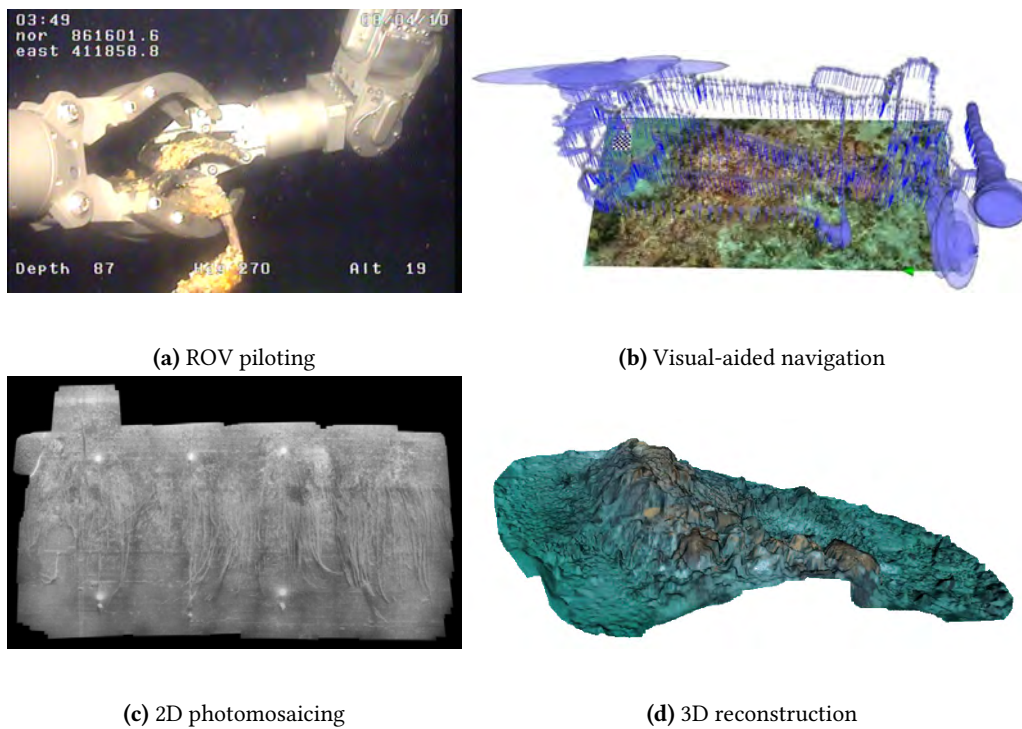
**I**N this chapter we present the main problem that has motivated this thesis: mapping of underwater environments under low visibility conditions. The motivations behind this problem are introduced in Section 1.1, relating them to the requirements of real-world applications and the limitations of current approaches. Next, we state the objectives of the thesis in Section 1.2 and we briefly describe, in Section 1.3, the context in which this work has been carried out. Finally, the organization of the thesis document is presented in Section 1.4.



## 1.1 Motivation

Over the past few years, unmanned underwater vehicles have greatly improved as a tool for undersea exploration, inspection and intervention. Remotely Operated Vehicles (ROVs) have reduced the need for manned submersibles, increasing the safety and the duration of underwater operations. This type of vehicles are nowadays routinely used in offshore industries as well as in science applications. However, the need for a large and expensive support vessel to be able to handle the crane, the tether and the involved ROV crew and operators has stimulated research towards the development of Autonomous Underwater Vehicles (AUVs). Free from the limitation of a physical connection to a surface ship, AUVs allow for extended operations at lower costs, providing stand-alone platforms that can gather data close to the seafloor without human supervision and avoiding the risks associated to the umbilical cable. Nonetheless, the autonomy of these vehicles presents challenges on multiple levels, beginning with its autonomous navigation and localization. Since ubiquitous absolute measurements such as those provided by Global Positioning System (GPS) are not available underwater, AUV navigation and localization must rely on other solutions. A range of different possibilities has been explored throughout the years: arrays of acoustic beacons deployed on the seafloor; dead reckoning using a combination of depth, inertial and Doppler velocity sensors; terrain based navigation against an *a priori* known map or Simultaneous Localization And Mapping (SLAM) approaches that allow to concurrently build a map of the vehicle's environment and use it to obtain estimates of its location. The research conducted during past decades on these and other areas such as vehicle control, path planning or mission planning has led to a successful use of AUVs in many applications including marine geology [Yoerger et al.1998, Kelley et al.2005, German et al.2008, Paduan et al.2009], marine biology, [Armstrong et al.2006, Williams et al.2010], underwater archaeology [Foley et al.2009, Bingham et al.2010], fisheries management [Clarke et al.2009], under-ice exploration [Kunz et al.2009], disaster response [Camilli et al.2010] or even, more recently, intervention tasks [Ribas et al.2012].

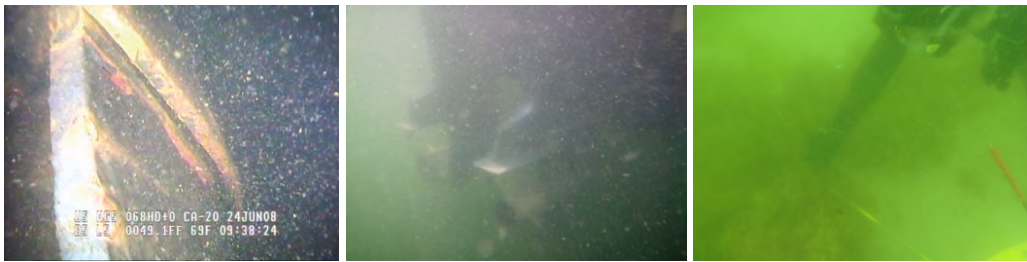
Integral to all these unmanned platforms, image acquisition plays a major role to most underwater applications (see Figure 1.1). Virtually all missions performed with ROVs require some form of visual data gathering, not only for the sake of observing and collecting information but to provide visual feedback to the pilot. In regard to AUVs, navigation, localization, and mapping through optical imaging have been key topics for researchers in both underwater robotics and marine science communities. Underwater navigation has benefited from video imagery in visual odometry and visual SLAM approaches [Gracias et al.2003, Eustice et al.2008, Salvi et al.2008, Wirth et al.2013] to provide drift-free navigation by using motion estimates obtained from onboard camera frames. Optical imagery has been also used to construct two-dimensional (2D) underwater photomosaics that enable the overview of extended areas through the registration of individually collected images. 2D photomosaics have become a key tool to locate and map areas or objects of interest, detect changes or plan subsequent missions in an area with applications in dam in-



**Figure 1.1:** Example of procedures requiring the underwater acquisition of visual data.

spection [Ridao et al.2010], marine geology [Escartín et al.2008], underwater archaeology [Singh et al.2004, Bingham et al.2010], environmental monitoring [Elibol et al.2011b] and damage assessment [Lirman et al.2010]. Furthermore, in scenarios with a high three-dimensional (3D) component, optical data has also been used to model 3D reconstructions of relevant underwater structures through structure from motion approaches using monocular video sequences [Pizarro et al.2004, Nicosevici et al.2009] or through the deployment of stereo camera systems [Zhang and Negahdaripour2010, Johnson-Roberson et al.2010].

Though unquestionably useful, the acquisition of optical images in the underwater domain is not without challenges. Light attenuation, water turbidity, and the inability to get within proximity of the target are but a few of the difficulties. It is well known that propagation of light underwater suffers from exponential attenuation [Duntley1963] and thus sunlight can only penetrate to a limited extent. Therefore, the inclusion of artificial lighting systems becomes a requirement to acquire optical images when operating in non-shallow areas. Although today's vehicles are equipped with increasingly improved optical cameras and powerful lighting systems, visibility is still often compromised due to water turbidity. In the presence of suspended particles, light wavelengths, which are small compared to the particle sizes, are blocked, deflected and scattered yielding poor or no visibility of the underlying sea floor (see Figure 1.2). This becomes a critical issue in a significant number of underwater surveying and mapping tasks that are carried out in turbid waters and murky environments. Mapping of harbors, dam inspections, ship-hull inspections or monitoring of bays and estuaries are a few applications typically conducted under



**Figure 1.2:** Different examples of optical images affected by backscatter and/or turbid water conditions.

poor visibility conditions. In those scenarios vehicles equipped only with optical systems (i.e. cameras or lasers) are heavily constrained by their limited visibility range. Furthermore, when a vehicle needs to perform an inspection close to the seabed, the action of its thrusters may quickly stir up suspended sediments which can also affect the quality of the acquired images. In the case of ROVs, the operator is forced to wait for the sediment to settle down and get back a video visual, which slows down the overall inspection task and consequently increases the associated costs.

Knowing the limitations of optical devices, underwater operations have long relied on sonar technology. Acoustic waves are significantly less affected by water attenuation, facilitating operation at greater ranges and allowing work in turbidity conditions thanks to larger wavelengths. Thus, sonar devices address the main shortcomings of optical sensors though at the expense of providing, in general, noisy data of lower resolution and more difficult interpretation.

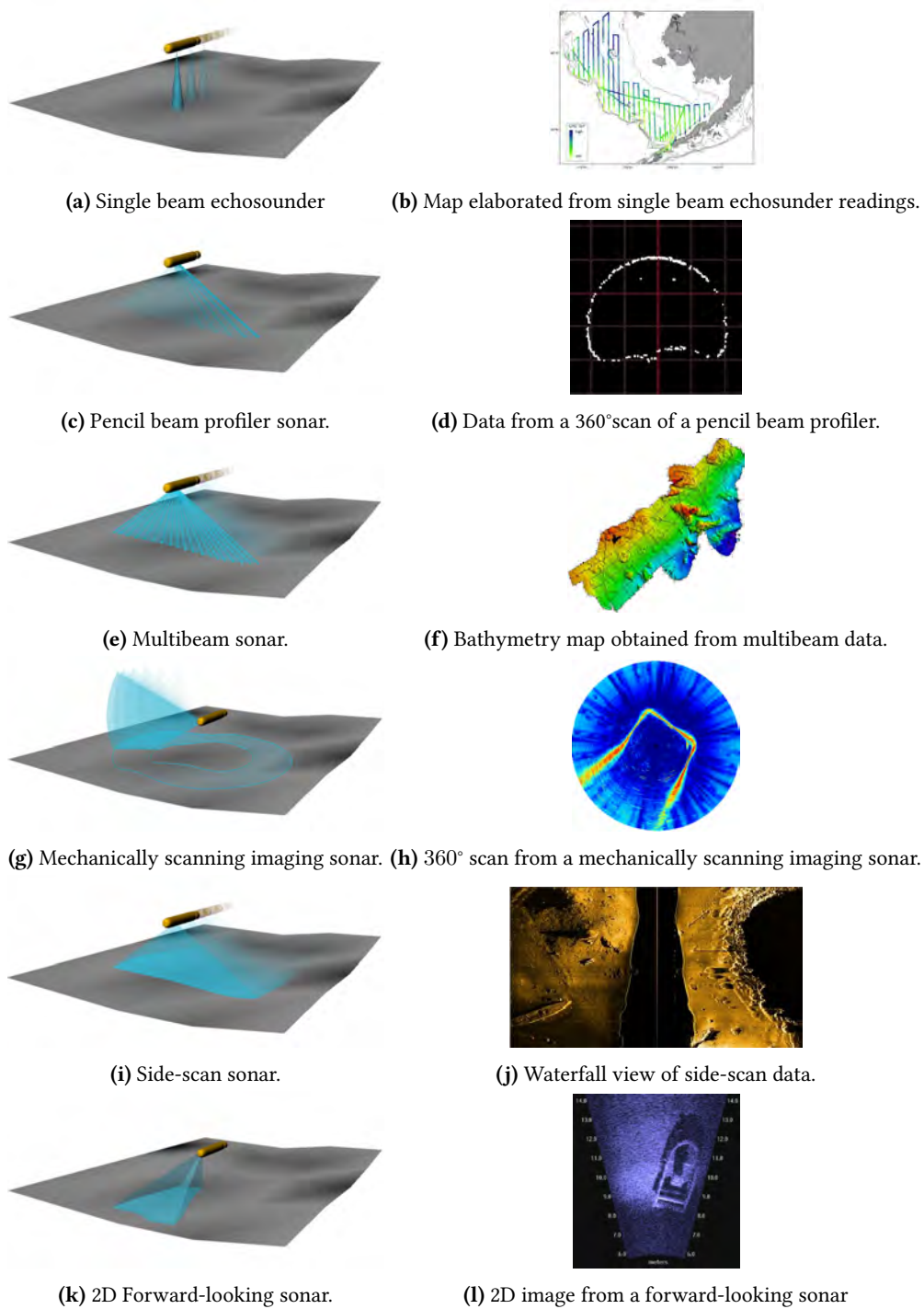
Sonars delivering range measurements, such as single beam echosounders (Figure 1.3a), profiling sonars (Figure 1.3c) or multibeam echosounders (Figure 1.3e), have been successfully employed for obstacle avoidance, navigation, localization and mapping [Leonard et al.1998, Tena et al.2003, Roman and Singh2005, Fairfield et al.2007, Kinsey et al.2006], the latter being especially popular for the creation of seafloor bathymetric charts. Imaging sonars, such as mechanically scanning imaging sonars (Figure 1.3g) or side-scan sonars (Figure 1.3i), have also been widely used in obstacle-avoidance, localization and particularly in mapping applications [Mallios et al.2014b, Ribas et al.2008, Tena Ruiz et al.2003, Aulinas et al.2010] thanks to its ability to represent the returning acoustic intensities from an insonified area.

Recently, a new generation of imaging sonars [Soundmetrics Corp.2013, BlueView Technologies Inc.2013, Tritech Gemini2013], namely the two-dimensional Forward-Looking Sonars (FLSs) (Figure 1.3k) are emerging as a strong alternative for those environments with reduced visibility given their capabilities of delivering high quality acoustic images at a near-video frame rate. FLS provide significant advantages over other imaging sonars, thanks to the use of advanced transducer arrays that allow simultaneous sampling of multiple acoustic returns and render them in a 2D image. For instance, mechanically scanning sonars require the rotation of an acoustic beam through different steps to gather the returning reflections along an angular sector. Therefore, the acquisition of a scan image involves a relatively long time and introduces distortions as a

consequence of the vehicle motion during the scan. Likewise, when using side-scan sonars, the seabed reflections of the two side beams are recorded in a series of cross-track slices that must be stitched along the motion direction in order to generate an overview of the scene and facilitate its interpretation. Hence, side-scan sonars are also subject to geometric distortions due to different instabilities of the deploying platform, thus requiring a correction either by using navigational and attitude information or through statistical techniques applied directly on the scan lines [Cobra et al.1992]. FLSs, on the other hand, directly deliver 2D acoustic images, which provide a closer rendition of what the eye naturally sees and thus minimize the required level of processing and interpretation.

Hence, FLS can be seen as an analogous tool of optical cameras in turbid waters. Immediate applicability is found in ROV operations, where FLSs are a valuable tool towards more efficient maneuvers in low to zero visibility. However, beyond that, applications are numerous and diverse. Being an excellent tool to observe and identify targets in poor visibility, FLS imagery has been employed in monitoring of fish populations [Baumgartner and Wales2006], inspection of underwater structures [Chen et al.2011] and automated detection of targets on the seafloor [Galceran et al.2012].

Several researchers have also drawn attention to the use of FLS either as a substitute or as a complementary device for optical cameras in sonar-aided navigation or mapping applications. The integration of FLS data in a visual SLAM framework to constrain the navigation drift of AUVs has been tackled within the context of ship hull inspection [Walter2008, Johannsson et al.2010, Hover et al.2012]. Even more straightforward is the sensor parallelism for mapping purposes: FLS can be exploited to mosaic the seafloor through the registration of FLS frames, following the same concept of 2D photomosaicing. Even though the range of FLS is greater than that of optical cameras, their Field of View (FOV) is also limited. Thus, it is often not possible to image a target area within a single frame or at least to do so without sacrificing a great deal of resolution by pushing the device's range to the limit. In such circumstances, mosaicing of FLS images allows obtaining an extended overview of an area of interest regardless of the visibility conditions and without compromising the resolution. However, despite the problem analogy, it must be noted that the particularities of FLS imagery, such as low resolution, low Signal-to-Noise Ratio (SNR) and intensity alterations due to viewpoint changes, pose a significant challenge to the techniques typically used in photomosaicing. Figure 1.4 illustrates the typical workflow to create an image mosaic. The process usually starts by computing frame to frame transformations using an image registration method. Then, consecutive images are aligned by transforming them to a common reference frame through compounding of the different transformations. Small misalignment errors that accumulate along the trajectory can be corrected by means of global optimization techniques that make use of the transformations between non-consecutive images. Finally, the images are blended together in a smooth mosaic. Each of these steps, and in particular image registration, is heavily conditioned by the inherent differences between optical and acoustic cues

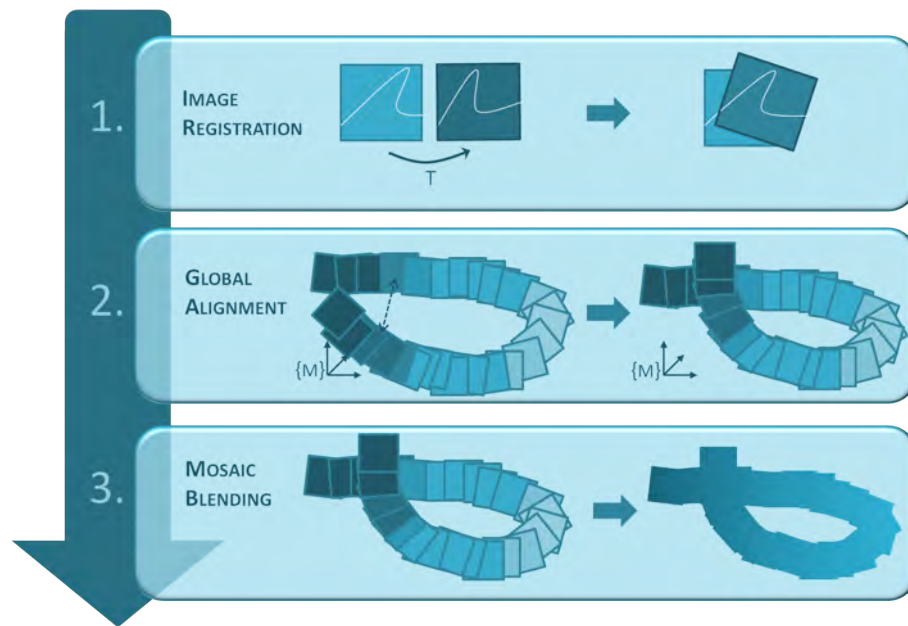


**Figure 1.3:** Different underwater sonar types and its data.

and therefore issues arise in trying to leverage the techniques used on optical images.

These difficulties, together with the still recent proliferation of high-resolution FLS devices on the market and their high cost when compared to other sonars, are perhaps to be blamed for the reduced efforts on FLS mosaicing. During the last decade, the specific problem of mosaicing



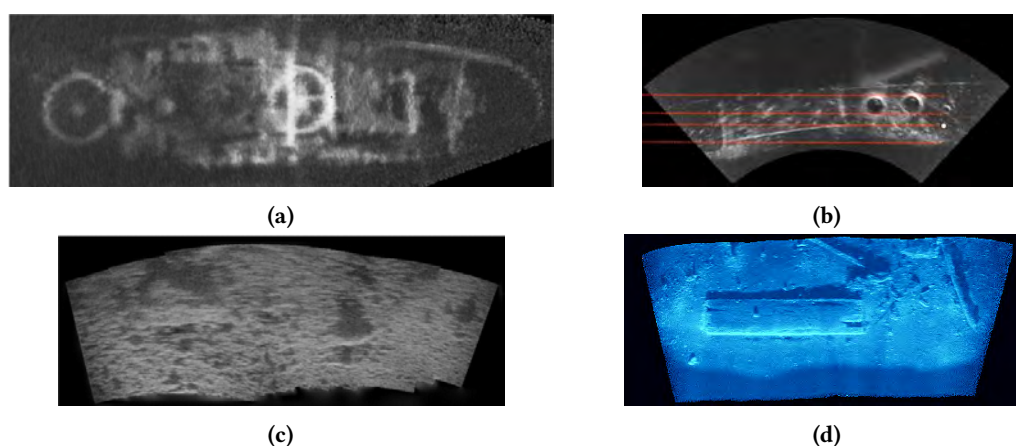


**Figure 1.4:** Typical mosaicing pipeline.

FLS imagery has only been tackled by a handful of researchers [Kim et al.2005, Negahdaripour et al.2005, Negahdaripour et al.2011]. Related state of the art includes also few works dealing with FLS image registration [Johannsson et al.2010, Aykin and Negahdaripour2012], which, as introduced, is an integral step in mosaicing. However, the field is still remarkably incipient and the existing mosaicing approaches have shown very limited results in terms of scale and complexity. Although it would be desirable to create consistent mosaics undergoing both translational and rotational 2D motions, extending along various vehicle tracklines, and applicable in a wide variety of environments, most of the reported mosaics in the literature are restricted to only a few frames gathered in a single trackline while imaging feature-rich scenarios (see Figure 1.5).

On the commercial side, some companies have also envisaged the potential of the FLS mosaicing capability. A spin-off company [AcousticView2014] which arose from previous research in the field [Kim et al.2005] is providing an initial mosaicing software for DIDSON sonar images [Sound Metrics DIDSON2013] suffering also from the aforementioned limitations. Some FLS manufacturers [BlueView Technologies Inc.2013] and other third party companies [Oceanic Imaging Consultants, Inc.2014] are providing software to mosaic FLS data in real-time through the use of absolute positioning sensors. These systems are conceived to work with FLSs deployed from surface boats so that GPS measurements can be used as the underlying positions to project the acoustic images and obtain a mosaic composition. Hence, they do not perform registration of FLS images and their results totally depend on the availability and accuracy of an absolute positioning system, thus not being a good option for underwater vehicles.

Therefore, the development of a full, versatile, and efficient mosaicing pipeline for FLS images would clearly push the envelope of vehicle mapping capabilities in low-visibility underwater



**Figure 1.5:** Example of FLS mosaics reported in previous works. (a) Shipwreck mosaic elaborated from 38 DIDSON sonar images, in [Kim et al.2005]. (b) Mosaic from few DIDSON images shown in [Negahdaripour et al.2011]. (c) Mosaic of 16 DIDSON images, in [Negahdaripour et al.2005]. (d) Small mosaic of an archaeological site, from [AcousticView2014].

environments. Moreover, it would contribute to the progress of other challenging subjects such as sonar-aided navigation or, in general, to any application that may benefit from context-awareness in poor visibility conditions.

## 1.2 Objectives

Once the motivations have been set forth, we state the overall goal of this thesis as follows:

*To develop a complete mosaicing pipeline, tailored to the characteristics of forward-looking sonar imagery, for the generation of acoustic mosaics that provide a consistent overview of approximately-planar underwater environments regardless of the visibility conditions.*

This general goal can be broken down to the following more specific objectives:

**Registration method for FLS images:** To propose a 2D registration methodology for the robust alignment of FLS frames coping with the challenging characteristics of this type of imagery, dealing with translational and rotational motions, being amenable to be implemented in real-time and applicable to a wide variety of environments.

**Integration inside a global alignment framework:** To develop a front-end to integrate the proposed registration method into a global alignment framework in order to achieve a consistent set of image poses based on the registration of consecutive and non-consecutive frames along the vehicle trajectory. We aim to enable the mosaicing capability without the requirement of absolute positioning sensors and, when possible, with minimally instrumented vehicles.

**Rendering of smooth acoustic mosaics:** To propose a blending strategy to render visually pleas-

ant and informative acoustic mosaics by correcting the photometrical irregularities that may be encountered when fusing FLS frames and provide a final composition with enhanced SNR and resolution with respect to the individual sonar frames.

**Real-time mosaicing:** To design and implement each of the pipeline steps in an efficient way, enabling the online construction of maps that can contribute to applications requiring vehicle (or pilot) situational awareness.

**Experimental validation with real data:** To validate the proposed methodology with multiple experiments in the context of relevant field applications and using data gathered through different FLS devices.

### 1.3 Context

The research of this thesis has been conducted at the Underwater Robotics Research Center, *Centre d'Investigació en Robòtica Submarina* (CIRS) of the Computer Vision and Robotics (ViCOROB) Institute of the University of Girona. Research in underwater robotics has been ongoing there since 1992, supported by several Spanish and European research programs. The group has developed several AUV prototypes: GARBI [Amat et al.1999], a vehicle originally conceived as a ROV that was restyled as an AUV; URIS [Batlle et al.2005], a lightweight AUV; Ictineu [Ribas et al.2007], which won the first Student Autonomous Underwater Challenge - Europe (SAUC-E) competition in 2006; Sparus, which championed SAUC-E in 2010 and has recently been restyled and presented as a commercial platform [Carreras et al.2013]; and GIRONA 500 [Ribas et al.2012], a reconfigurable AUV that has been used in the experimental part of this thesis. Research at CIRS revolves around AUV applications, and has focused on control architectures [Ridao et al.2002] [Palomeras et al.2012], model identification [Carreras et al.2003], machine learning [Carreras et al.2001, El-Fakdi and Carreras2013], mission control [Palomeras et al.2006], AUV intervention [Prats et al.2012b, Ribas et al.2012], SLAM [Ribas et al.2008, Mallios et al.2014b] and path planning [Hernández et al.2011, Galceran Yebenes2014]. Together with recent efforts in bathymetry mapping [Zandara et al.2013] and sonar scan matching [Mallios et al.2014a], this thesis represents one of the first endeavors of the group in acoustic mapping, opening not only a new research direction but providing new navigation and mapping capabilities to the CIRS vehicles.

Of special relevance to this thesis is also the work developed at the Underwater Vision Lab (UVL) from ViCOROB Institute. Outstanding research on underwater computer vision has been carried out at UVL during recent years, covering topics such as large-scale underwater mosaicing [Ferrer et al.2007], global alignment techniques for optical mapping [Elibol et al.2011a], image blending [Prados et al.2014], 3D scene modelling [Nicosevici et al.2009] and image-based classification [Shihavuddin et al.2013]. The joint work between the two labs has lead to successful results in real-world applications such as dam inspection [Ridao et al.2010], AUV mapping of



archaeological sites [Gracias et al.2013], or AUV mapping and intervention in a harbor scenario [Prats et al.2012a], which are operations in demanding environments that have stimulated as well the work presented in this doctoral dissertation.

Moreover, the work presented herein has contributed to the following projects in which CIRS has participated:

- CICYT Project AQUAVISION: Vision Systems for computer cartography and underwater aquaculture (Ref DPI2007-66796-C03-02), from which this thesis was supported through the FPI grant BES-2008-006095, funded by the Spanish Ministry of Education and Science.
- FP7 EU Project TRIDENT: Marine robots and dexterous manipulation for enabling autonomous underwater multipurpose manipulation (Ref FP7-ICT-2009-248497), funded by the European Commission.
- FP7-7 EU Project PANDORA: Persistent Autonomy through Learning, Adaptation, Observation and Re-planning (Ref FP7-ICT-2011-7-288273), funded by the European Commission.
- MINECO Project RAIMON: Autonomous Underwater Robot for Marine Fish Farms Inspection and Monitoring (Ref CTM2011-29691-C02-02), funded by the Spanish Ministry of Science and Innovation.
- MINECO Project COMAROB: Robótica cooperativa Marina para el mapeo acústico y la intervención (Ref DPI2011-27977-C03-02), funded by the Spanish Ministry of Science and Innovation.

Finally, this thesis has also benefited from a research stay in the Ocean Systems Lab (OSL) of Heriot-Watt University (Edinburgh, UK). OSL is a science and engineering research centre distinguished for its innovations in autonomous systems and underwater acoustic sensors theory and processing [Petres et al.2007, Petillot et al.2001, Tena Ruiz et al.2004, Reed et al.2004, Coiras et al.2007].

## 1.4 Document Roadmap

The remainder of this doctoral dissertation is organized as follows (see Figure 1.6). **Chapter 2** describes the principle of operation of a FLS together with the geometry model that describes its image formation and the challenges involved in processing FLS imagery. The information compiled in this chapter is the basis of FLS imaging and is essential to contextualize the methods and procedures presented in the following chapters.

The next three chapters contain the main stages of the proposed FLS mosaicing pipeline.

**Chapter 3** addresses the pairwise registration of overlapping FLS images. After reviewing the state-of-the-art of FLS registration techniques, we propose a Fourier-based technique to register FLS images in a robust and efficient way. The different steps of the method are thoroughly detailed,

emphasizing the particularities that have been adapted to cope with the noise and artifacts of acoustic images. Finally, in order to validate the method's performance we present quantitative comparisons against other state of the art registration techniques.

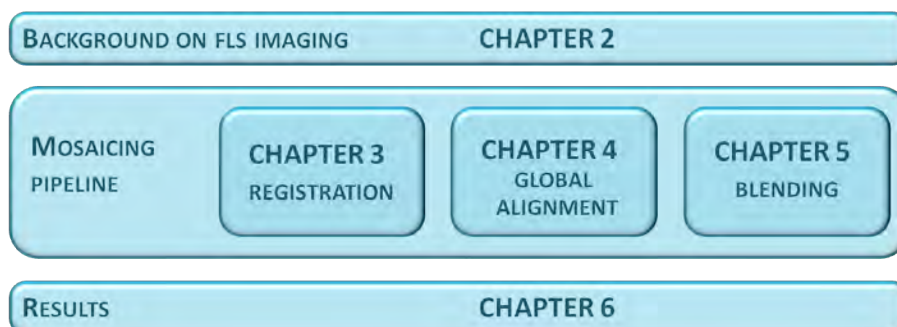
**Chapter 4** deals with the global alignment stage by means of a pose-based graph optimization. We describe the pose constraints that are introduced to the graph and we propose a method to estimate the uncertainty of a pairwise registration in order to appropriately weight these constraints in the optimization. We also present the workflow to construct the graph, involving the candidate selection criteria that is followed to attempt frame registrations along the vehicle trajectory, for both offline and online mosaicing.

**Chapter 5** covers the rendering of the individual FLS frames into a smooth, visually-pleasant and informative mosaic. An analysis of the differences between the optical and acoustic image blending problem sets the basis for the adopted approach. We identify the photometric irregularities that might be encountered in FLS mosaicing and we present a series of strategies to diminish their impact in the final mosaic. Moreover, we discuss upon other rendering-related issues such as the mosaic's SNR improvement and the super-resolution possibilities.

**Chapter 6** presents the experiments and results that validate the full mosaicing pipeline. Experiments with real datasets including relevant field applications such as ship hull inspection, harbor mapping or the exploration of an archaeological shipwreck are presented and analysed.

Lastly, the document is concluded in **Chapter 7** summarizing the main contributions and the future research directions to be explored.

Two accompanying appendices are found at the end of the main thesis body. **Appendix A** is a recap, from a user point of view, of good practices and general guidelines to collect FLS imagery in a way that facilitates its posterior mosaicing. **Appendix B** encloses an example of object detection algorithm on FLS images in the context of an underwater chain inspection scenario. The algorithm takes advantage of the mosaicing capability to perform detection on FLS images with a higher degree of reliability.



**Figure 1.6:** Chapter organization outline.



# 2

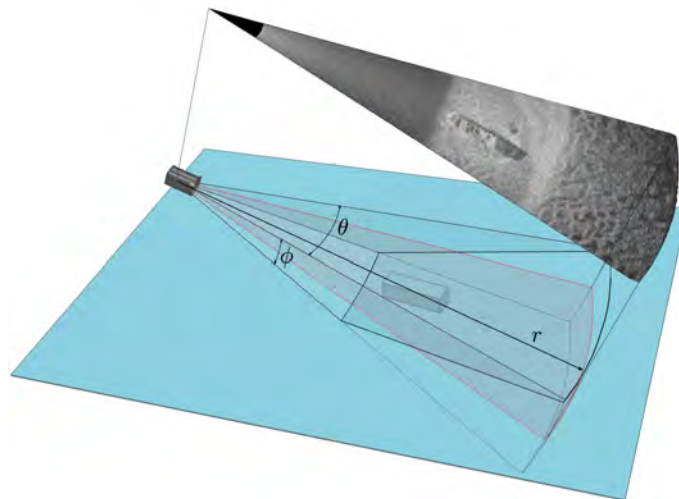
## **BACKGROUND ON FORWARD-LOOKING SONAR IMAGERY**

---

Before moving on to the different stages of the mosaicing pipeline, it is essential to understand the FLS image formation and establish a suitable model to describe the imaging geometry of the sonar. In this chapter, we first describe the operational principles of FLS devices (Section 2.1) by explaining the basics behind its acoustic image formation. Section 2.2 provides a description of the FLS geometry models used in the related state of the art, followed by a discussion on our model choice and its limitations. We also provide a summary of the main challenges to be faced when dealing with FLS imagery to better understand how they can influence subsequent processes such as image registration or blending.

## 2.1 FLS operation

Two-D 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. The sonar insonifies the scene with an acoustic wave, spanning its FOV in the azimuth ( $\theta$ ) and elevation ( $\phi$ ) directions, and then the intensity of the acoustic return is sampled by an array of transducers as a function of range and bearing (Figure 2.1). 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. Therefore, the 3D information is lost in the projection into a 2D image.



**Figure 2.1:** FLS operation. 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).

Following the nature of the transducer readings, the images are arranged and represented in polar coordinates. Therefore the dimensions of a raw frame correspond to the number of beams in the angular direction and the number of range samples in the range axes. This representation is then converted to the final 2D image in Cartesian coordinates for an easier interpretation. It is worth noting that this process produces images with non-uniform resolution as one pixel in polar domain is mapped to multiple pixels with the same intensity in the Cartesian coordinates as the range increases.

Beyond this operation principle, FLSs are being manufactured in a variety of specifications. There are different internal beamforming techniques, from electronic beamformers, that use frequency-steered arrays, to lenses beamformers using a set of lenses to focus the acoustic beams.

The use of lens simplifies the electronics and reduces power consumption though it requires more space and possible calibration procedures [Negahdaripour et al.2005] in order to rectify the lens distortion and ensure data to model consistency. Across the different available models we can find the well-known trade-off between operating frequency and range as well as a variety of FOV and beam angle widths. To give a flavor of typical numbers, table 2.1 summarizes the specifications of the main FLS models found in the market nowadays. During the development of this thesis we have worked with data from the first five listed devices.

## 2.2 Imaging geometry model

In this section we analyse the imaging geometry of FLSs and identify the transformation that relates two acoustic images under different models.

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:

$$\mathbf{P} = \begin{bmatrix} X_s \\ Y_s \\ Z_s \end{bmatrix} = \begin{bmatrix} r \cos \theta \cos \phi \\ r \sin \theta \cos \phi \\ r \sin \phi \end{bmatrix} \quad (2.1)$$

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

$$\mathbf{p} = \begin{bmatrix} x_s \\ y_s \end{bmatrix} = \begin{bmatrix} r \cos \theta \\ r \sin \theta \end{bmatrix} = \frac{1}{\cos \phi} \begin{bmatrix} X_s \\ Y_s \end{bmatrix} \quad (2.2)$$

As can be seen in Equation 2.2, 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, based on whether the 3D points are approximated through an orthogonal projection or not.

### 2.2.1 Orthographic projection approximation

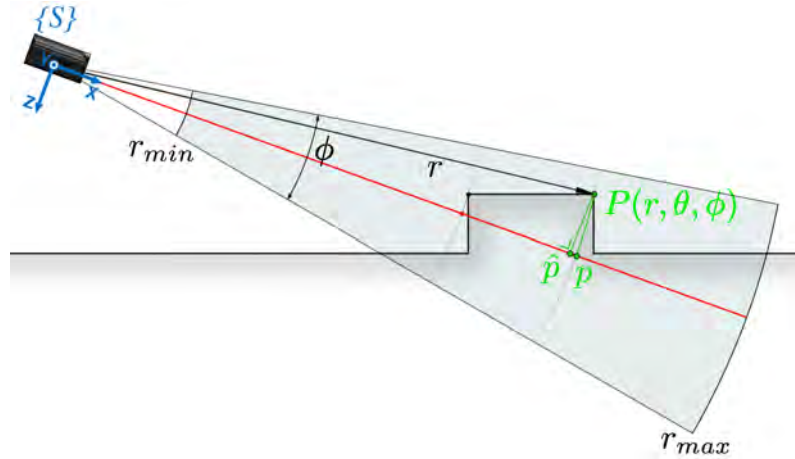
Given the narrow elevation angle that typically characterizes FLS devices (around 7-10 deg), the nonlinear component defined by  $\phi$  is tightly bound. Approximating this narrow elevation to the limit (i.e., considering only the zero-elevation plane), we end up with a linear model in which the sonar can be seen as an orthographic camera [Walter2008]. Hence, the projection  $\mathbf{p}$  of a 3D point  $\mathbf{P}$  is approximated by the orthogonal projection  $\hat{\mathbf{p}}$  as shown in Figure 2.2.

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

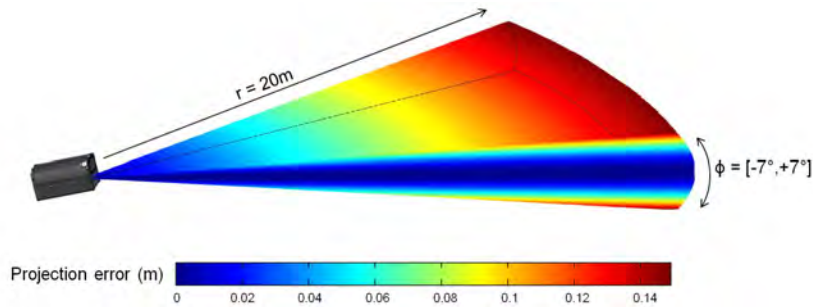
<b>SONAR MODEL</b>	<b>OPERATING FREQUENCY</b>	<b>RANGE (m)</b>	<b>FOV (deg)</b>	<b>NUM.OF BEAMS</b>	<b>BEAM WIDTH (<math>\theta \times \phi</math>) (deg)</b>	<b>MAX.UPDATE RATE (Hz)</b>	<b>MAX.RANGE RESOLUTION (m)</b>	<b>ANGULAR RESOLUTION (deg)</b>
<b>ARIS Explorer 3000</b>	1.8/3 MHz	0.5-20	30°	128	0.25°×14°	15	0.003	0.25°
<b>BlueView P900-130</b>	900 kHz	1-100	130°	768	1°×20°	15	0.025	0.18°
<b>BlueView P900-45</b>	900 kHz	1-100	45°	256	1°×20°	15	0.025	0.18
<b>DIDSON</b>	1.1/1.8 MHz	0.4-40	29°	96	0.3°×14°	21	0.0025	0.3°
<b>Tritech Gemini i720</b>	720 KHz	0.2-120	120°	256	1°×20°	30	0.008	0.5°
<b>Imagenex 965</b>	260 KHz	0.5-300	120°	480	1.5°×20°	10	0.0025	0.25°
<b>Kongsberg M3</b>	500 KHz	0.2-150	140°	256	0.95°×30°	40	0.01	0.95°
<b>R2Sonar Sonic2024</b>	200-400 KHz	1-400	160°	256	1°×20°	60	0.0125	0.5°
<b>RESON Seabat 7128</b>	200/400 KHz	1-500	128°	256	0.54°×31°	50	0.025	0.5°

**Table 2.1:** Specifications of main FLS devices present in the market during the elaboration of this thesis.

by the projection approximation is a function of the distance in the  $X_s Y_s$  plane and the vertical distance to the point [Johannsson et al.2010] (see Figure 2.3). 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.



**Figure 2.2:** Sonar projection geometry. A 3D point  $\mathbf{P}(r, \theta, \phi)$  is mapped into a point  $p$  on the image plane along the arc defined by the elevation angle. Considering an orthographic approximation, the point  $\mathbf{P}$  is mapped into  $\hat{p}$ , which is equivalent to consider that all scene points rest on the plane  $X_s Y_s$  (in red).

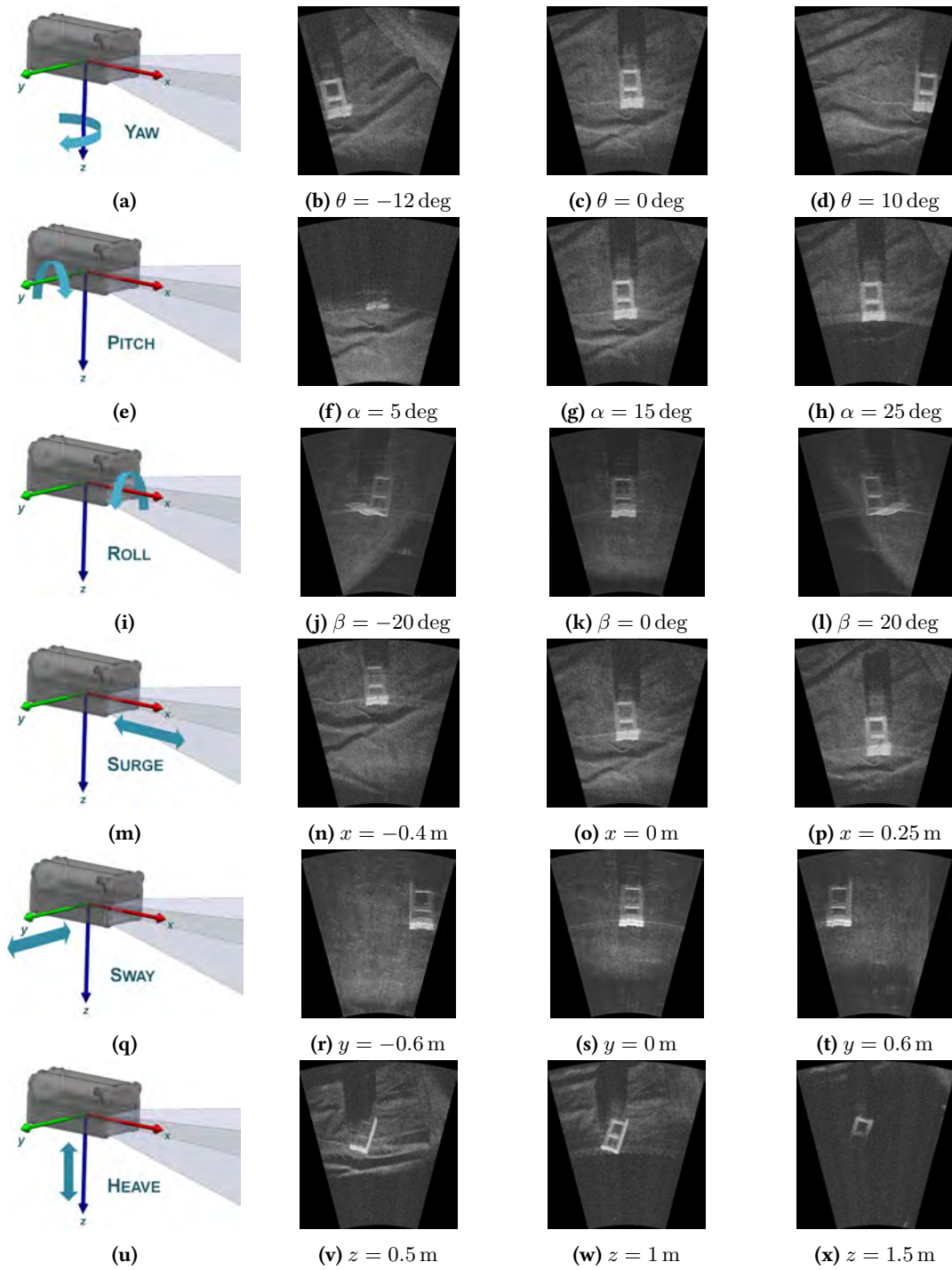


**Figure 2.3:** Projection error due to the orthographic approximation. Example computed for a sonar with a range up to 20 m and an elevation angle of 14 deg. As can be observed, the error increases with the range and extreme elevation angles. However, thanks to the narrow elevation angle of the FLS sonars, the projection error is bounded within few centimetres for typical parameters.

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 (DOF). Finally, notice also that changes in  $z$  will only change the insonified area and its intensities but will not have any impact in the scale of the images as long as the device's range is not changed.

Therefore, by using this model, a point in the space represented by  $\mathbf{p}$  and  $\mathbf{p}'$  in two different





**Figure 2.4:** Effect of rotations and translations on FLS images. Each row shows an example sequence under the motion indicated in the first column.

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 ( $\theta$ ):

$$\mathbf{p}' = \mathbf{H}\mathbf{p} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & t_x \\ \sin(\theta) & \cos(\theta) & t_y \\ 0 & 0 & 1 \end{bmatrix} \mathbf{p} \quad (2.3)$$

### 2.2.2 Non-approximated projection

Other approaches in the literature [Sekhati and Negahdaripour2007, Aykin and Negahdaripour2012] 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 [Negahdaripour2012b]:

$$\mathbf{p}' = \mathbf{H}\mathbf{p} = \begin{bmatrix} \gamma q_{11} & \gamma q_{12} & \psi q_{13} \\ \gamma q_{21} & \gamma q_{22} & \psi q_{23} \\ 0 & 0 & 1 \end{bmatrix} \mathbf{p} \quad (2.4)$$

where  $\gamma = \cos \phi / \cos \phi'$ ,  $\psi = r \sin \phi / \cos \phi'$  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 non-uniform 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 [Negahdaripour2012a], has also been used in the context of 3D sonar motion estimation [Aykin and Negahdaripour2013]. In theory, it allows handling all 6 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 [Negahdaripour2012a].

However, these models require the knowledge of the elevation angles at every image location, which are 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 as shown in [Negahdaripour2012a]. The estimation is obtained using the sonar pitch, the sonar height (provided by an altimeter) and assuming that the points located at the leading and trailing edges of the images correspond to the minimum and maximum set ranges (as shown in Figure 2.5a) and therefore they are located at the extreme  $\phi_{min}$  and  $\phi_{max}$  values. Moreover, Aykin and Negahdaripour [Aykin and Negahdaripour2012] have shown that this estimation of the elevation map can be defined with higher accuracy by incorporating the elevation angles of prominent features. To that end, they make use of object-shadow pairs extracted from detected

blob regions to estimate the elevation angle of 3D features.

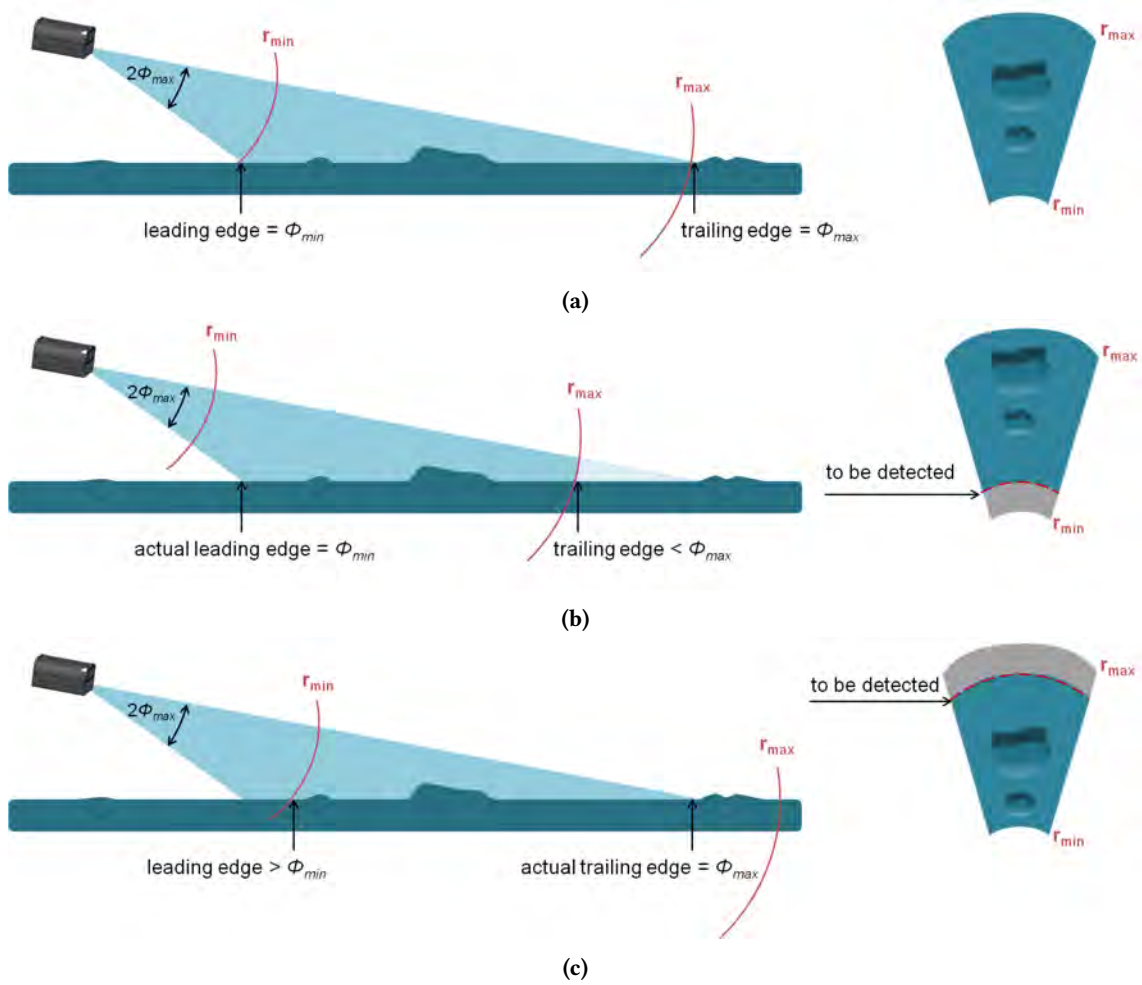
### 2.2.3 Selecting a FLS geometry model for 2D mosaicing

The use of the exact geometry model provides two main advantages when compared to the approximated version (Equation 2.3). On one hand it allows the estimation of the sonar motion in the vertical direction ( $z$ ). However, for 2D mapping purposes this estimate is not required and in case the sonar motion were to be estimated, the translations in the vertical direction can 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.

The other advantage comes from the incorporation of the elevation angles in the registration process, which reduces the errors introduced by the orthographic approximation and has proved to enhance the local image alignment [Negahdaripour2012a]. Nevertheless, this improvement 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. There are many cases that the elevation angles cannot be reliably estimated from the imaging configuration as proposed in [Negahdaripour2012a]. Although imaging parameters can be adjusted, it is difficult that in a real scenario the image limits correspond to minimum and maximum set ranges (see Figure 2.5). Moreover, the detection of the actual leading and trailing edges on the images in order to establish this point-elevation correspondence can be complex and inaccurate. On the other hand, we cannot rely on a robust shadow detection to derive the elevation angles of image features as shown in [Aykin and Negahdaripour2012]. The distinctiveness of a shadow area is subject to intensity patterns that depend on the topology and the reflectivity of the imaged environment. Therefore, shadow detection can become ambiguous and introduce errors in the estimation of the elevation angles (see Figure 2.6).

All in all, we chose to adopt the simplified 2D model, though being an approximation, it is suitable to describe the image formation process and set the basis for the subsequent registration process. Moreover, 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. As it will be argued in Chapter 3 this is of special interest for acoustic images where the inclusion of all the image information will minimize ambiguities in the registration.

It is worth emphasizing that the main limitation of the FLS imaging geometry, namely the assumption of the imaged scene being nearly planar, can be relaxed thanks to the range extent of the FLSs, which can vary from tens to hundreds of meters depending on the device. 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

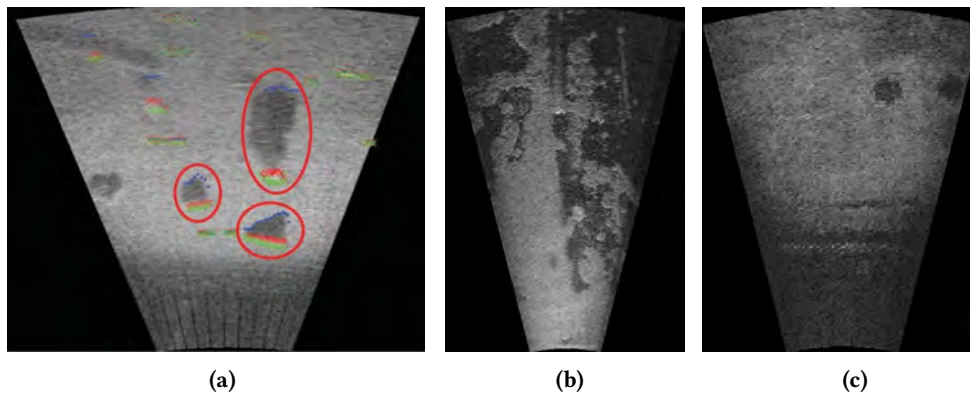


**Figure 2.5:** Elevation angle estimation from leading and trailing image edges. (a) Ideal case where the set ranges correspond to the leading and trailing image edges and have the extreme values  $\phi_{min}$  and  $\phi_{max}$ . (b) Example of a trailing edge not corresponding to  $\phi_{max}$  and minimum range not corresponding to the real leading edge of the image. In this situation, the real leading edge could be detected in the image. (c) Example of a leading edge with a  $\phi$  angle bigger than the  $\phi_{min}$  that is assumed for the minimum range. The actual trailing edge is on  $\phi_{max}$  but its location would have to be detected in the image as it does not correspond to the maximum range.

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) could be considered to accommodate the imaging configuration so as to match the planar assumption as closely as possible.

## 2.3 Challenges in FLS imagery

As introduced in Chapter 1, 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



**Figure 2.6:** Feature shadow detection for elevation estimation. (a) For certain isolated features it is easy to detect and measure shadow cues to estimate its elevation angle. Example extracted from [Aykin and Negahdaripour2012]. However shadow detection can be ambiguous in many environments, for instance in surfaces with mixed reflectivity (b) or in the presence of holes (c).

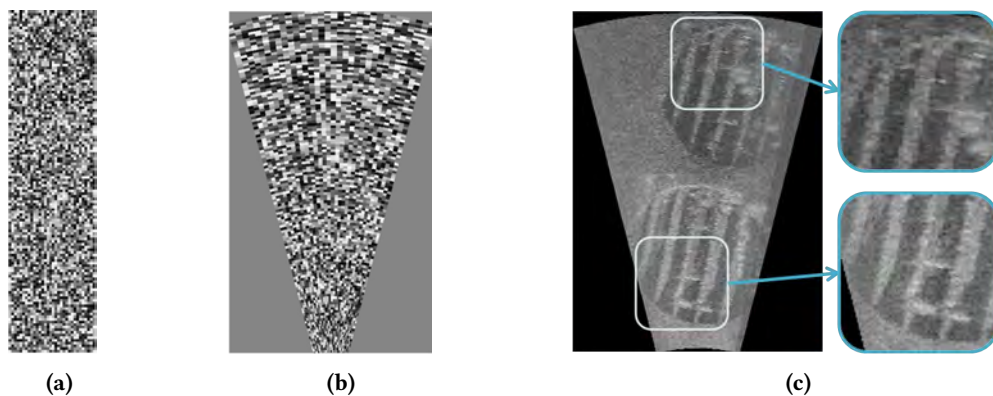
closely related to the nature of sonar image formation that may impact subsequent processing steps. In the following we provide a summary of the most relevant ones.

- **Low resolution:** Although they are considered high resolution 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. The cross-range or angular resolution is limited by the number of transducers that can be physically packed in the device while the down-range one depends on the operating transducer frequency. Both cross-range and down-range resolutions are important for image quality and discriminating between closely spaced objects/targets. As can be seen on table 2.1, the highest resolution sonar nowadays [Sound Metrics ARIS2013] is capable of delivering a down-range resolution of 3 mm and a cross-range around 3 cm at 10 m range. However most of the devices have larger beam widths and lower frequencies that translate to lower resolutions. Moreover, as a consequence of the sensor's polar nature, measurement sparseness increases with the range when represented in a Cartesian space (see Figure 2.7). For instance, for the ARIS sonar configured at 3 MHz where the minimum range is 1 m and the maximum range is 10 m, the ratio between the largest and the smallest pixel size can be up to 10 which means that a pixel in the polar image can occupy from 1 to 10 pixels in the Cartesian image. Therefore, as mentioned before, this results in a non-uniform resolution that contributes to degrade the visual appearance of the image.

- **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. As will be seen in Chapter 5, this SNR can be enhanced by registering and averaging multiple frames of the same scene. Therefore mosaicing intrinsically provides a way to alleviate this problem.





**Figure 2.7:** Inhomogeneous resolution in FLS images. (a) Example of a random polar frame. (b) Corresponding Cartesian frame where the effect of the domain transformation can be clearly appreciated (pixels being more sparsely mapped as the range increases). (c) Example image where content located at farther ranges exhibits lower resolution.

- **Inhomogeneous insonification:**

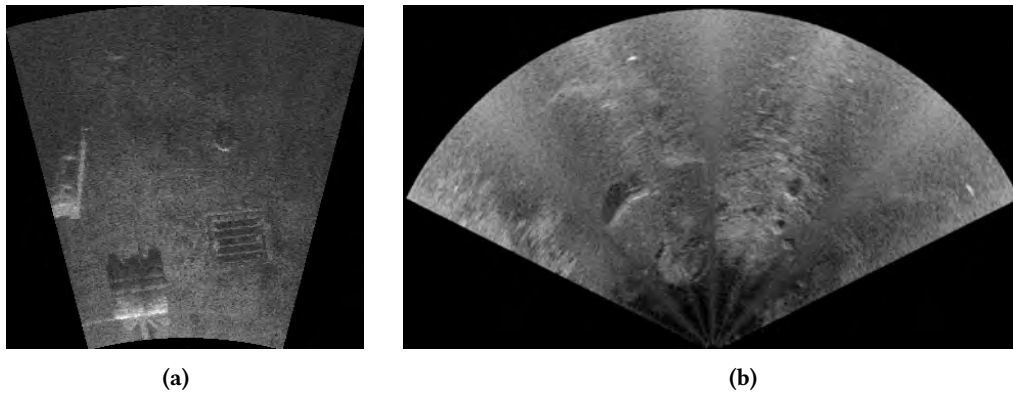
FLS devices typically include a Time Varying Gain (TVG) mechanism that compensates for the transmission loss (i.e., the spreading and/or attenuation of the sound pulse on its travel to the target and back). In this way, similar targets located at different ranges are perceived with similar intensities. However, due to the imaging geometry, a change on the FLS incidence angle and/or the inclination of the imaged surface can introduce variations in the image illumination either within a sequence or within a single frame (see Figure 2.8a). Besides, other inhomogeneous intensity patterns can arise due to differing sensitivity of the lens or transducers according to their position in the sonar's FOV [Negahdaripour et al.2005] (see Figure 2.8b). When strongly present, these inhomogeneous intensities can affect the image registration step and they obviously have an impact on the image blending step as well. As will be described in Chapter 5, these artifacts can be alleviated by estimating the inhomogeneous intensity pattern from the averaging of a sufficient number of images or through some image equalization or regularization method [Kim et al.2005].

- **Viewpoint changes:**

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.

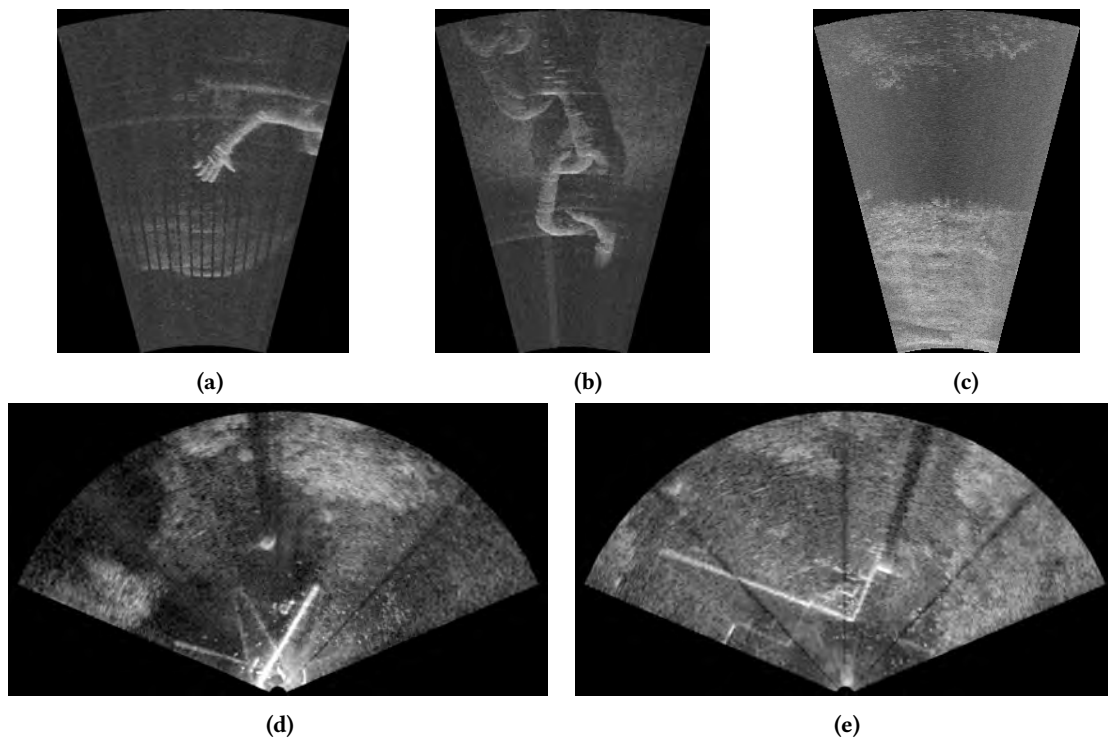
- **Other artifacts:**

Under some circumstances, spurious content can appear in the sonar images causing ambiguity



**Figure 2.8:** Inhomogeneous insonification in FLS images. (a) Example of inhomogeneous insonification in a planar terrain due to sonar tilt angle. (b) Inhomogeneous insonification due to the overlapping of transducer beams across the FOV.

in the registration: reverberation artifacts, acoustic returns from the water surface, cross-talk between beams, or multipath that generates ghosting artifacts. However, these artifacts can generally be minimized by adopting a proper configuration and imaging setup.



**Figure 2.9:** Other artifacts present in FLS images. (a) Aliasing of tank reverberation overlaid in the image as black stripes on the background. (b) Crosstalk. Bright returns of the chain links repeated in the sidelobes of the main beam resulting in bright arcs. (c) Top part of the image exhibiting water surface reflections. (d-e) Ghosting effects of main structures due to multipath.

# 3

## REGISTRATION OF FORWARD-LOOKING SONAR IMAGES

---

**T**HE pairwise registration of frames is a key step in mosaicing. Given a pair of images gathered from different viewpoints it allows to find the planar transformation that properly aligns them into a single and common reference frame. The computer vision community has proposed numerous registration methods over the last few decades [Zitova and Flusser2003], however most of them have been designed to deal with optical images. As seen in the previous chapter, FLS images are affected by different problems, tightly related to their mode of operation. Therefore, there is the need to find a suitable technique for performing FLS registration. In this chapter we start by analyzing the different methods that are employed in previous works to register FLS imagery. We then propose a registration technique that can overcome their limitations and cope with the challenging characteristics of the acoustic data. Finally, we provide quantitative comparisons of the proposed method to testify its performance against two other registration techniques used in earlier works.



## 3.1 Related work on FLS Image Registration

### 3.1.1 Feature-based registration at pixel level

Feature-based methods are arguably the most popular type of registration methods existing in the literature. They are based on establishing correspondences between a limited set of well localized and individually distinguishable points (i.e., features) detected in each of the images [Tuytelaars and Mikolajczyk2008]. The traditional pipeline for feature-based registration is sketched in Figure 3.1. The process begins with the detection of image features or salient points that are relevant according to content of their surroundings. In order to later establish correct correspondences across the images, these points need to be repeatable, so they can be located again in other images, and discriminative, so they can be correctly identified. The feature detection is followed by a description process that computes feature descriptors, i.e., a compact representation of the neighborhood of a feature. Descriptors can be a simple fixed-size window of pixels around the feature point or more complex descriptor vectors that characterize the gradients of the surrounding texture. After that, there is a matching step where the point-to-point correspondences from the two images are established. To that end, similarity measures are evaluated on the descriptors of the two images to rank the goodness of a particular match. Typical approaches use the similarity between statistics computed in a window around the feature point (e.g., normalized cross-correlation or sum of squared differences) or the similarity between feature descriptors vectors (through some distance measure). In order to avoid wrong correspondences, this process is usually followed by an outlier rejection step that discards wrong matches using a robust estimation algorithm (e.g., RANdom SAmple Consensus (RANSAC) [Fischler and Bolles1981]). Finally, the matches that are deemed as correct (i.e., inliers) are used to estimate the transformation that relates the images through orthogonal regression.

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 Harris corner detector [Harris and Stephens1988] and matched by searching over small local windows. Similarly, in the work of Kim *et al.* [Kim et al.2005, Kim et al.2006], Harris features extracted at the third and fourth level of a Gaussian pyramid scale are matched with cross-correlation of local patches. Once correspondences are established, the transformation estimation is performed with a RANSAC-alike method that uses the least square error between correlations as a criterion for determining outliers. Negahdaripour *et al.* [Negahdaripour et al.2011] highlight the complexities of mosaicing benthic habitats with FLS images and show the difficulty of registering DIDSON frames from a natural environment by using the popular Scale-Invariant Feature Transform (SIFT) detector and descriptor [Lowe2004]. Results report a very low percentage of inliers in the matching step (about 8%) and only small displacements of features located in the shortest ranges of the images could be effectively matched.

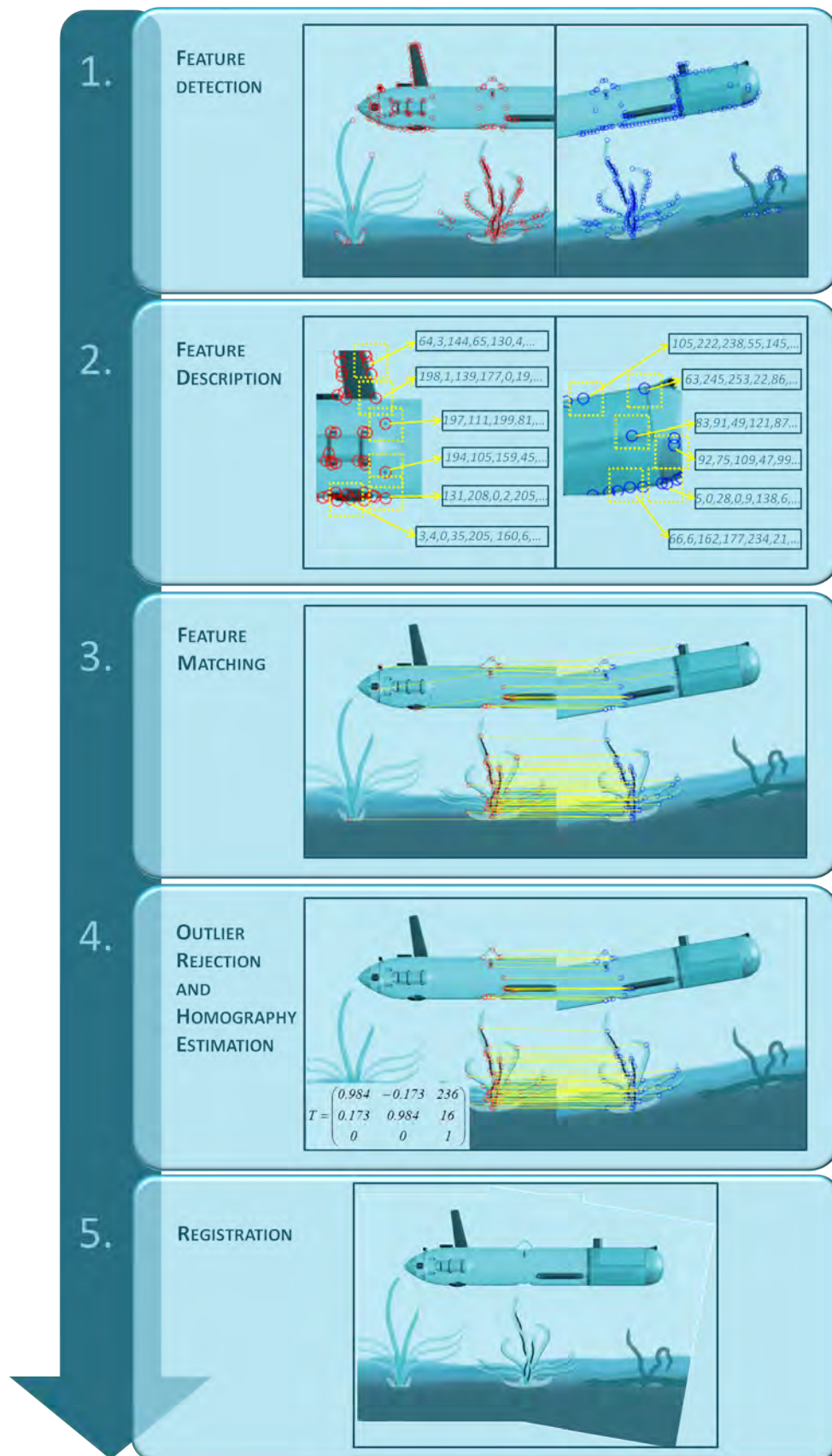


Figure 3.1: Feature-based registration pipeline.

In general, due to the inherent characteristics of sonar data, pixel-level features extracted in sonar images suffer from low repeatability rates [Hurtós et al.2013] as it will be shown later in Section 3.3. Consequently, extracted features lack stability and are prone to originate erroneous matches and yield wrong transformation estimations. 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.

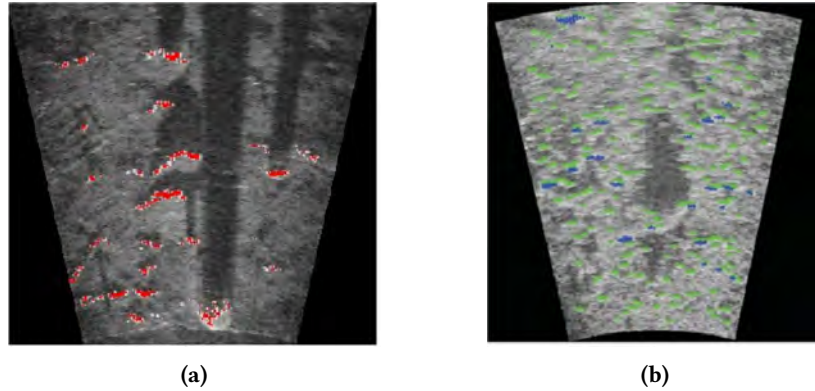
### 3.1.2 Feature-based registration at region level

The unreliable behavior of the pixel-level features over acoustic images has not gone unnoticed by other researchers who have proposed alternatives involving features at region level rather than at pixel scale. Johannsson *et al.* [Johannsson et al.2010] proposed the extraction of features in local regions located on sharp intensity transitions (i.e., changes from strong to low signal returns as in the boundaries of object-shadow transitions). In their approach the sonar images are first smoothed with a median filter, then their gradients are computed, and points exceeding a given threshold are finally clustered into features (see Figure 3.2a). 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 and Straßer2003]. 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. The same feature-based registration has been later used in [Hover et al.2012].

A similar approach has been recently presented by Aykin and Negahdaripour [Aykin and Negahdaripour2012, Aykin and Negahdaripour2013]. Instead of thresholding on the gradient domain, blob features are clustered from two types of pixels in the image: on one hand, the highest intensity values (i.e., brighter pixels assumed to be objects or structures on the ground surface) and on the other, those pixels with a negative vertical gradient that ensure object-shadow transitions and not viceversa (see Figure 3.2b). As an alternative to the NDT algorithm, Aykin and Negahdaripour propose the use of an adaptive scheme where a Gaussian distribution is fitted to each blob feature. Afterwards, an optimization is formulated to seek the motion that best fits the blob projections from one Gaussian map to the other.

### 3.1.3 Taking it a step further: area-based registration

Following the trend observed in the state of the art, 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 expect to be able to handle more



**Figure 3.2:** Feature-based registration at region level. (a) Example of extracted features from [Hover et al.2012]. (b) Example of extracted features from [Aykin and Negahdaripour2013]. Points with higher intensities are in blue and points with negative gradients in green.

changes in the visual appearance of the image and minimize the ambiguities in the registration. The common shortcoming of area-based techniques is their inability to handle complex transformations, being limited to the estimation of similarity transforms. However, and according to the simplified FLS geometry model that we adopted in Section 2.2.3, 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 existing area-based approaches, including spatial-correlation, Fourier-based methods and mutual information methods, we propose the use of Fourier-based techniques. The particularities of these kind of methods suggest that they can be appropriate for the registration of FLS imagery since, by design, they offer some robustness to noise, illumination changes and occlusions [Feroosh et al.2002] while being more computationally efficient than spatial correlation or mutual information approaches [Zitova and Flusser2003].

### 3.2 Fourier-based registration for FLS

Fourier transforms and frequency domain analysis are a cornerstone of signal processing. It is not surprising then that Fourier-based methods, and in particular the phase correlation algorithm [De Castro and Morandi1987,Reddy and Chatterji1996] that will be introduced hereafter, have been widely employed in many 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 to compute the Fourier transforms. In a similar problem to the one we tackle in this thesis, phase correlation has been applied to register underwater optical images in order to build photomosaics [Eustice et al.2002,Bülow et al.2009]. 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 use them to 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 [Chailloux2005, Vandrish et al.2011] while other researchers employed it in the registration of 2D and 3D sonar range scans [Bülow et al.2010, Bülow and Birk2011]. However, to the best of the author's knowledge, its use on FLS data has never been explored.

In the following we introduce the main principle of Fourier-based registrations for estimating translational shifts. Later, in section 3.2.2, we will see how this is extended to obtain also rotational estimates and finally we will provide the complete outline of the proposed registration method for FLS images.

### 3.2.1 Phase correlation for the estimation of translational shifts

The principle of the Fourier-based registration techniques resides in the Fourier shift property. This property establishes that a shift between two functions (e.g., images) is transformed in the Fourier domain into a linear phase shift. Let  $i_1(x, y)$  and  $i_2(x, y)$  be two images related by a 2D shift  $(t_x, t_y)$ , namely

$$i_1(x, y) = i_2(x - t_x, y - t_y). \quad (3.1)$$

Then their 2D Fourier transforms, denoted by  $I_1(u, v)$  and  $I_2(u, v)$ , are related as follows:

$$I_1(u, v) = I_2(u, v) \exp^{-j(ut_x + vt_y)}. \quad (3.2)$$

As can be seen in Equation 3.2, the phase information contains the shift of spectral content between the images, whereas the amplitude information will convey the amount of energy within a particular frequency.

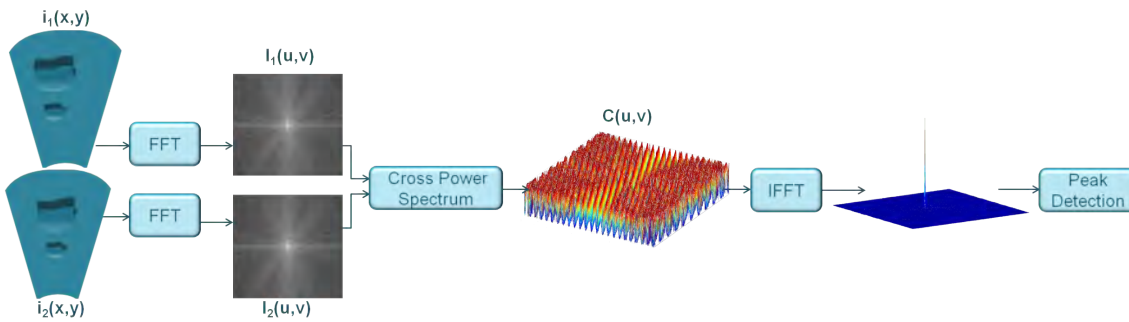
The phase term can be factored out by:

$$C(u, v) = \frac{I_1(u, v) I_2^*(u, v)}{|I_1(u, v) I_2^*(u, v)|} = \exp^{-j(ut_x + vt_y)} \quad (3.3)$$

where  $I^*$  denotes the complex conjugate of  $I$ . This can be regarded as a special type of correlation, in which the images cross-spectrum is normalized, so that the frequency components have all unity amplitude while retaining the original phase values, thus decoupling the signal energy from the structure. Hence, the normalizing denominator 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 [Froosh et al.2002].



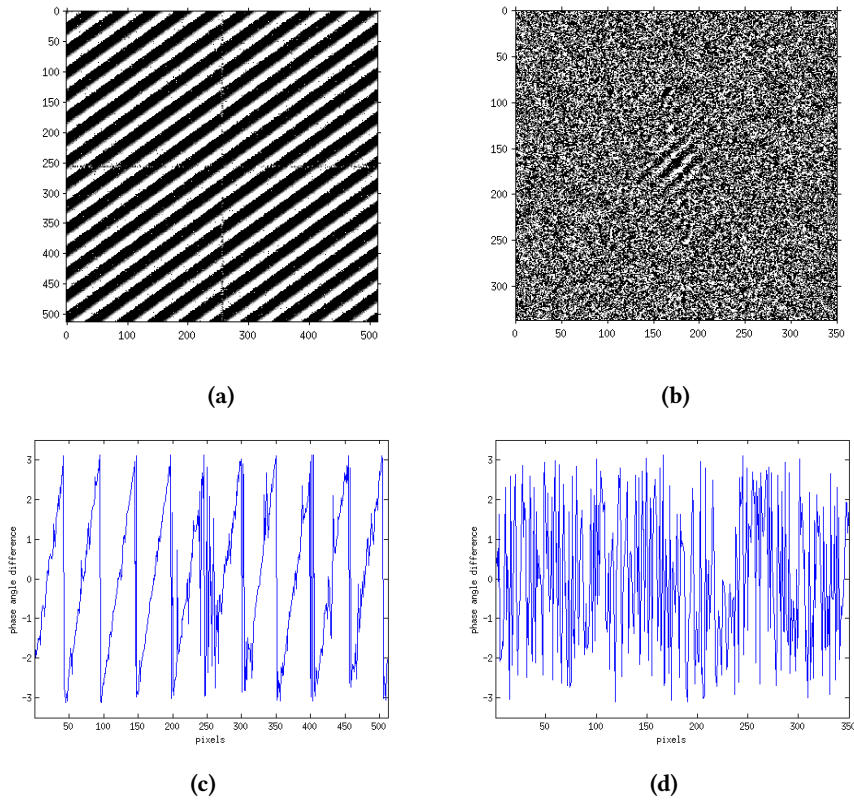
The most common way to solve Equation 3.3 for  $(t_x, t_y)$  is to apply the inverse Fourier transform to the phase correlation matrix  $C(u, v)$ . In the ideal case, this corresponds to a 2D impulse (Dirac function) centered on  $(t_x, t_y)$ , that directly leads to the identification of the integer 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, sub-pixel displacements can be estimated through the fitting of different functions on the vicinity of the peak [Foroosh et al.2002, Ren et al.2010].



**Figure 3.3:** Basic phase correlation workflow.

A different group of approaches, initiated by the work of Hoge *et al.* [Hoge2003], try to recover the offsets in Eq. 3.3 by working only in the frequency domain. Following their formulation, the image shifts can be computed as the slopes of a plane fitted to the phase difference data, which corresponds to the phase angle of  $C(u, v)$ . Computing these slopes implies the 2D unwrapping of the phase data which is a notoriously ill-posed problem [Hoge2003]. To avoid this step, Hoge *et al.* proposed to perform a subspace approximation of the noisy phase correlation matrix imposing Rank-1 constraint. In this way the 2D unwrapping problem is reshaped into two 1D unwrappings of the phase correlation dominant eigenvectors. Later on, Balci and Foroosh [Balci and Foroosh2006] have shown that actually no phase unwrapping is needed as the discrete phase difference matrix of two images corresponds to 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 several publications reporting successful results on optical images and Magnetic Resonance Imaging (MRI) [Hoge2003, Balci and Foroosh2006], the implementation of these methods does not seem feasible on FLS images. Figure 3.4 shows an example of the phase difference matrices obtained from two optical images (Figure 3.4a) and two FLS images (Figure 3.4b). While it is possible to compute reliably the width of a sawtooth cycle in the optical case, the cycles are hardly distinguishable in the FLS example. Even after attempting filtering operations on the phase difference matrix, the robust estimation of the offsets from the phase difference cycles is impracticable.

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



**Figure 3.4:** 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 (i.e., the shifts) in the sonar case.

the peak in the spatial domain, reason for which we developed on the standard approach.

### 3.2.1.1 Edge-effects

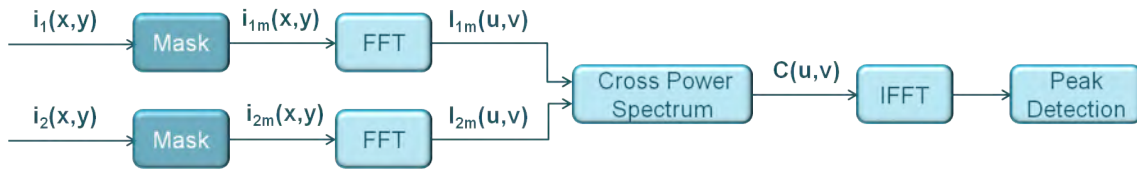
There are some factors not linked to the image nature itself, but to the computation process of the Fourier transform, that can overshadow the phase correlation matrix and lead to failure in detecting the peak if not handled properly. The most critical one, are the so-called edge effects. The phase correlation theory described in the previous section holds for periodic signals and continuous Fourier transforms. In the discrete case, the FFT algorithm is used to approach the infinite Fourier transform, by imposing a cyclic repetition of finite-length images by tiling them. Then, the abrupt transitions generated between the image edges when the images are tiled result in undesired frequency components appearing in the Fourier spectrum, which is also known as spectral leakage. These introduced frequency components may alter the subsequent computation of the phase correlation matrix and the overall registration result. In order to minimize the spectral leakage, it is typical to perform a windowing operation before the FFT computation. This is achieved by applying an apodization function (e.g., Blackman, Cosine, Gaussian, Hamming,

Hanning) that provides a smooth transition at the image boundaries.

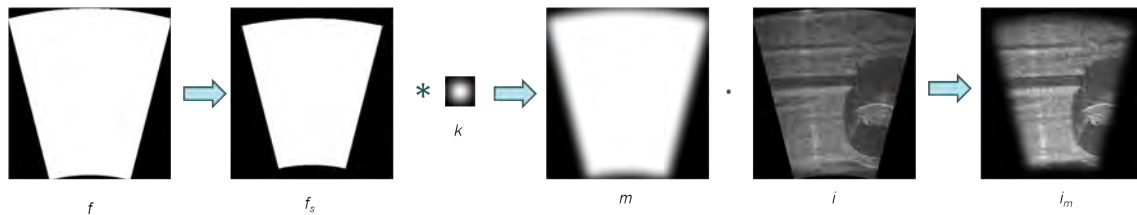
In a similar manner, the fan-shaped footprint edges of the FLS images in Cartesian coordinates introduce frequency components which are independent from the image content (see Figure 3.7e). Hence, the registration can latch and those components and not on the real image information. Back in the spatial domain, this causes a strong false peak around the origin of the matrix ( $t_x = 0$  and  $t_y = 0$ ) that can easily hide the location of the true peak. Thus, to avoid these effects in FLS images, we should apply a mask that tapers out the boundaries of the FLS Cartesian footprint prior to the FFT computation (see Figure 3.5). The image masking is performed through the following procedure, where  $n$  is a percentage of the maximum image size (typically around 3%):

1. Compute the footprint  $f$  of the input sonar image  $i$ .
2. Apply a shrink morphological operation to  $f$  by  $n$  pixels obtaining  $f_s$ .
3. Create a Gaussian filter  $k$  of size  $6n \times 6n$  and  $\sigma = n$ .
4. Create mask  $m$  convolving  $f_s * k$ .
5. Apply the mask to get the final masked image  $i_m = i \cdot m$ .

Note that this process will always result in the loss of some boundary image information. However, as shown in Figure 3.7, it is essential to preserve the spectrum of the image content and be able to detect the real shift.



**Figure 3.5:** Phase correlation workflow with masking. The block labelled as "Mask" applies the procedure described in Figure 3.6 to avoid edge effects.

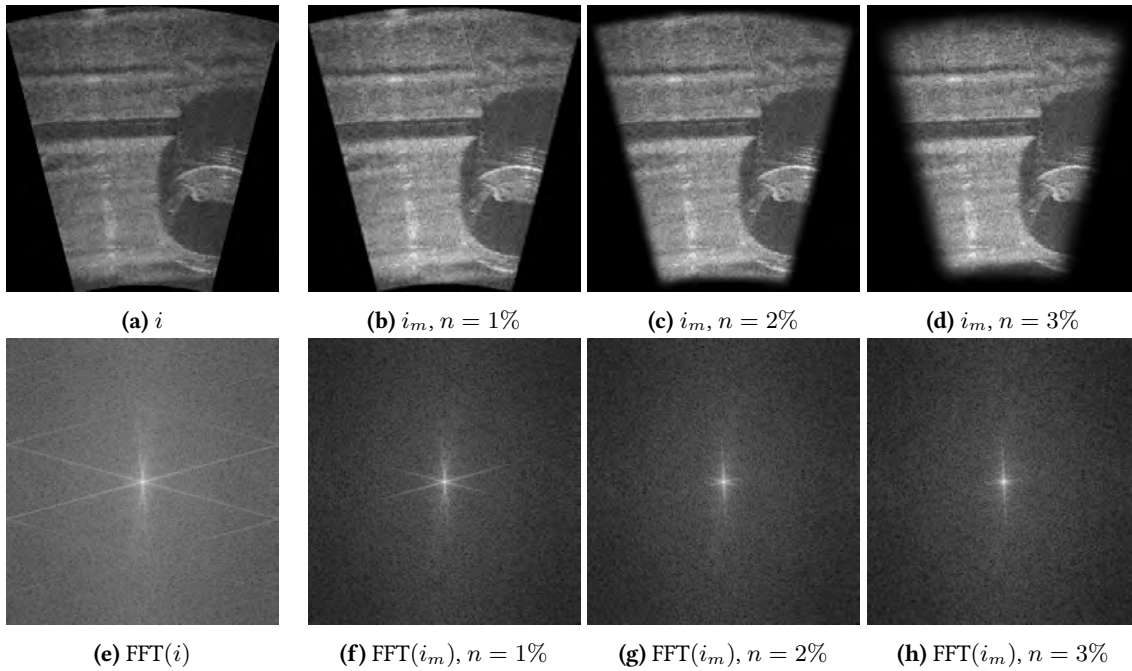


**Figure 3.6:** Image masking procedure.

### 3.2.1.2 Filtering operations in the phase correlation of FLS images

Besides the edge-effects, there are a number of factors that may result in arbitrary peaks when applying the inverse Fourier transform to the phase correlation matrix, thus reducing the ability of detecting a clear dominant peak.

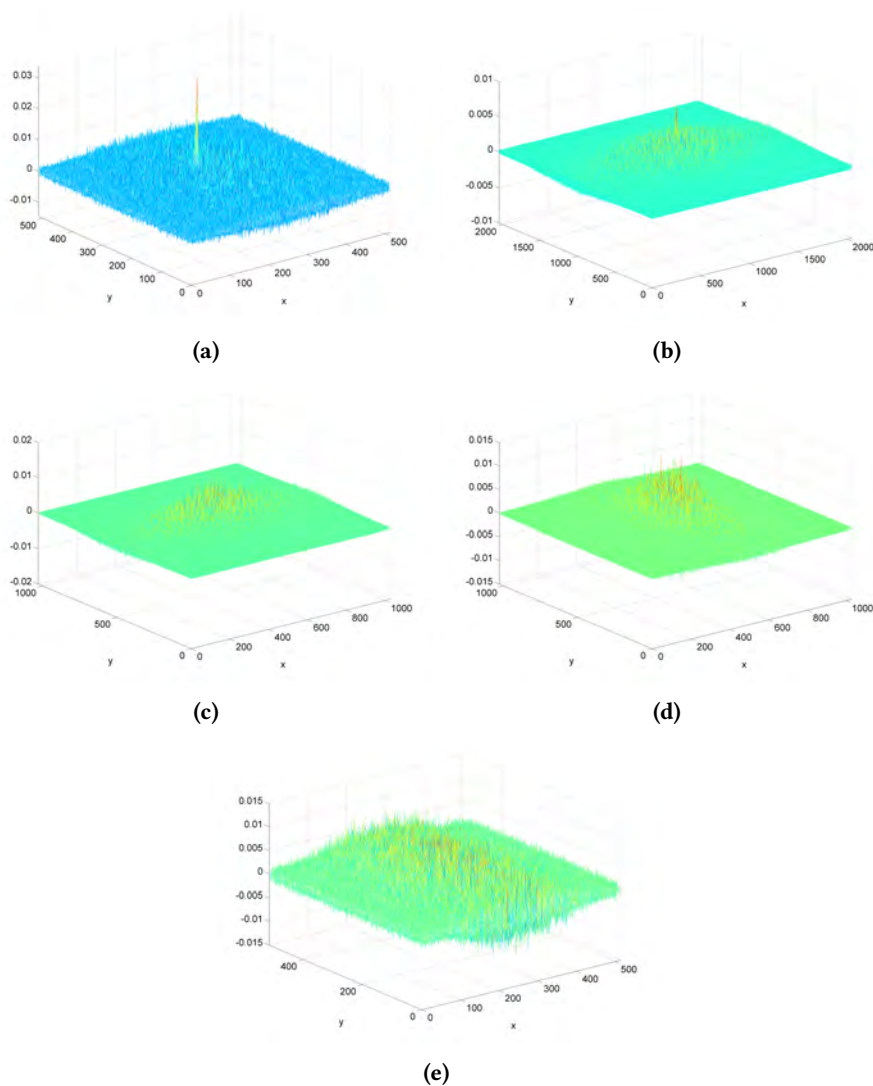




**Figure 3.7:** Masking of edge effects. The FFT (e) of the input image (a) exhibits strong patterns due to the footprint boundaries of the image (notice its perpendicularity to the image edges). By masking the image edges with a Gaussian kernel (b-d) the effect on the FFTs is alleviated (f-i). The smoother the transition, the higher the attenuation of the pattern present in the spectra. Notice how after diminishing the edge effects other components related to the image content gain relevance, such as the vertical edge due to the across-range shadow in the center of the image.

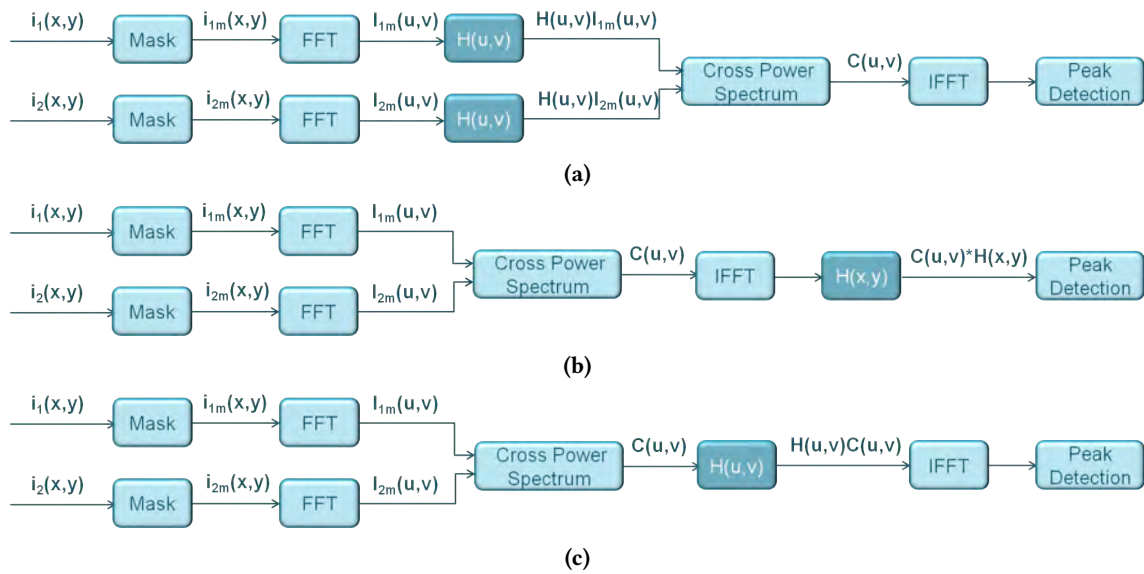
Due to the homogeneous treatment of all the frequency spectrum, phase correlation is highly sensitive to low SNR. Certain frequencies with very low SNR are given the same weighting as high signal content areas, thus leading to potential erroneous correlations and therefore spurious peaks. This is for instance the case of some high frequencies where an edge signal may be buried under speckle noise or the jagged pixel appearance product of the non-homogeneous resolution.

Other factors, such as intensity alterations caused by different viewpoints, errors introduced by the approximated geometry model or the image content of non-overlapping areas are likely to give rise also to multiple local maxima and reduce the amplitude of the true registration peak. Typically, the method itself is quite resilient to some of these noise sources, such as for instance the noise due to non-overlapping content. This is because the correlation signal (which can be approximated as a 2D Dirichlet distribution [Frooosh et al.2002]) is highly localized in the spatial domain while the Gaussian noise due to non-overlapping images is diffused, therefore being fairly easy to identify the peak even in low overlap conditions. However, as introduced in Section 2.3, FLS images are challenging and may often present a combination of all the abovesaid noise sources together, leading to a correlation surface with an unrecognizable main peak (see Figure 3.8). It becomes then fundamental to perform some filtering operations to attenuate the noise as much as possible and be able to recover the correlation peak with maximum reliability.



**Figure 3.8:** Examples of obtained peak surfaces when applying phase correlation to overlapping FLS images. Unfortunately, due to different noise sources it is often not possible to detect a clear dominant peak like in (a) or (b). Surfaces with scattered peaks (c-e) are common and compromise the detection of the true registration peak.

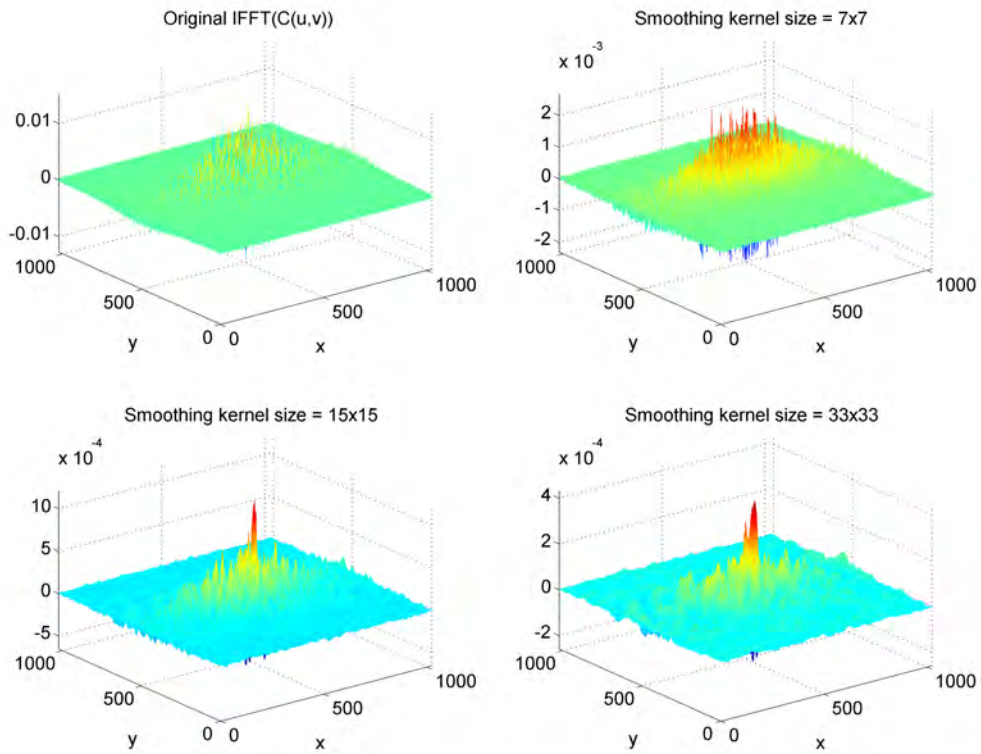
It is common practice to filter the image's spectra in order to attenuate unwanted frequencies that can lead to a noisy phase correlation matrix [Stone et al.2003] (see Figure 3.9a). However this requires a precise estimation or *a priori* knowledge of the spectra of the input images. Without such, determining these filters becomes critical as there is a risk of attenuating not only the unwanted components but also the discriminating phase components. Although the nature of the sonar images, as with most natural images, is of low-frequency, we cannot know beforehand the frequency responses that we might encounter when exploring a new environment. In general, we want to retain both low frequency characteristics, such as the reflectivity transition from a sandy area to vegetation, as well as high frequency components that arise from object edges or



**Figure 3.9:** Diagrams of prefiltering and postfiltering operations. (a) The Fourier transforms of the image are filtered before the correlation is computed. (b) The filtering takes place after the correlation, once back in the spatial domain. (c) The filtering takes place on the phase correlation matrix, in the frequency domain.

protruding seabed features. Therefore, in order not to limit the potential benefit arising from the correlation of these frequencies we chose to keep the input images unfiltered. Instead, we advocate for computing the cross power spectrum first and deal with the noise afterwards, once the phase of all frequencies has contributed (either for good or for bad) to the computation of the correlation matrix. From that point, two main approaches can be envisaged to deal with the noise: from the spatial (Figure 3.9b) or the frequency domain (Figure 3.9c).

**Filtering of the correlation surface in the spatial domain.** The most straightforward approach to minimize the noise effects is to apply a filter in the spatial domain, essentially enhancing the robustness of the peak detection. Once the phase correlation matrix is inverted, there will typically be several scattered peaks around the highest peak due to a combination of the aforementioned factors. Eventually, the vicinity of the highest peak contains peaks of smaller amplitude than the neighborhood of other peaks. The intuitive idea is that we should select as the highest correlation point the highest peak that is within a high neighborhood. A simple way to achieve this is by applying an averaging filter. Therefore before looking for the peak we can perform the convolution with an averaging kernel that will essentially smooth the noisy surface. However, choosing the right size for this kernel is critical and can severely condition the output of the registration: from selecting a totally wrong peak to losing a great deal of accuracy with a large filter that heavily smears the data (see Figure 3.10). Moreover, the appropriate size will vary depending on the topology of each obtained peak surface, and it is difficult to tune it in a precise and efficient way.

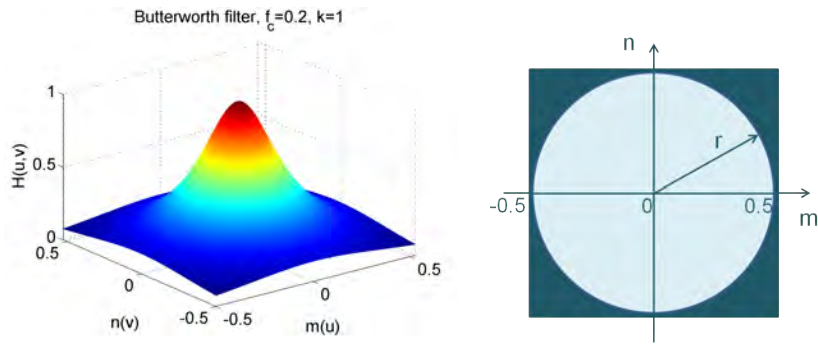


**Figure 3.10:** Effect of the averaging filter applied on the correlation surface for different kernel sizes.

**Frequency domain filtering of the phase correlation matrix.** Another approach can be devised by filtering in the frequency domain, before applying the inverse Fourier transform to the phase correlation matrix. The idea is to filter the noisy components of the phase correlation matrix so that they are not converted to spurious peaks in the spatial domain. Hence we will achieve a much cleaner surface where the final correlation peak can be detected in a more reliable manner. Moreover, the application of the filter in the frequency domain will be presumably faster than the convolution in the spatial domain for most of the kernel sizes. As stated earlier, the frequencies that are more likely to introduce problems are the high frequencies. Therefore, we start by considering the application of a low pass filter, such as a Butterworth filter (Figure 3.11). The Butterworth filter is a commonly used filter whose frequency response is described by:

$$H(u, v) = \frac{1}{1 + \left(\frac{r}{f_c}\right)^{2k}}, u = 1, 2, \dots, M - 1, v = 1, 2, \dots, N - 1 \quad (3.4)$$

where  $r = \sqrt{m^2 + n^2}$  is the frequency radius in normalized coordinates  $m = (u - M/2)/M$  and  $n = (v - N/2)/N$  within range  $[-0.5, 0.5]$ ,  $k$  is the filter order and  $f_c$  is the cutoff frequency in the range  $(0, 0.5]$ .



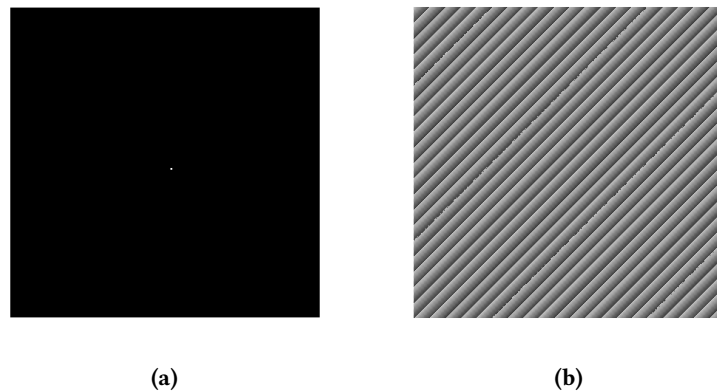
**Figure 3.11:** Butterworth filter.

Again, and analogously to the spatial case, we face the problem to tune the cutoff frequency of the filter. In this case, this is equivalent to finding the proper radius from the origin of the spectral plane in which the frequency components will begin to attenuate. In the literature there are few precedents suggesting the application of filters in order to mask the noisy components of the phase correlation matrix [Stone et al.2001, Eckstein and Vlachos2009, Pap et al.2012]. Stone *et al.* [Stone et al.2001] introduce a masking procedure to cut the noisy terms of the phase correlation matrix when estimating subpixel displacement, once integer pixel shifts have already been solved. They establish a fixed mask of radii  $r = 0.6(L/2)$  from the frequency origin, where  $L$  is the minimum number of samples in the  $x$  and  $y$  directions. Pap *et al.* [Pap et al.2012] adopt a brute-force style approach, performing a range of simulations to compute the shifts with a discretization of all possible filter radius to finally keep the best one. Eckstein and Vlachos [Eckstein and Vlachos2009] presented a robust phase correlation filter in the context of a particle image velocimetry application. Given the specific characteristics of the involved imagery they can accurately extract an analytical model of the SNR and therefore they can apply a Gaussian filter optimal for the spectral characteristics of the problem at hand.

Despite the brute-force approach, that we discard for obvious efficiency reasons, the other solutions are not versatile enough for our purposes. We are facing a more open problem, where we should deal with images that undergo a different amount of pixel shifts, that might have significantly different spectral characteristics depending on the observed content and different noise artifacts depending on the situation. Therefore, the phase correlation matrices that we can encounter can present substantial changes, and establishing a fixed filter cannot be regarded as a good solution. A too high cutoff may result in the effect of the unwanted correlations not being minimized. On the other hand, a too low cutoff can end up with a diluted peak centred at the origin, due to the correlation of the lowest frequency components such as the average intensity image (DC term). Hence, we propose an adaptive solution that adjusts the filter cutoff according to noise present in the phase correlation matrix. The key point is to automatically determine the cutoff frequency so that it minimizes those components that will lead to spurious spikes in the spatial domain, while preserving the location of the true delta of the correlation. Although



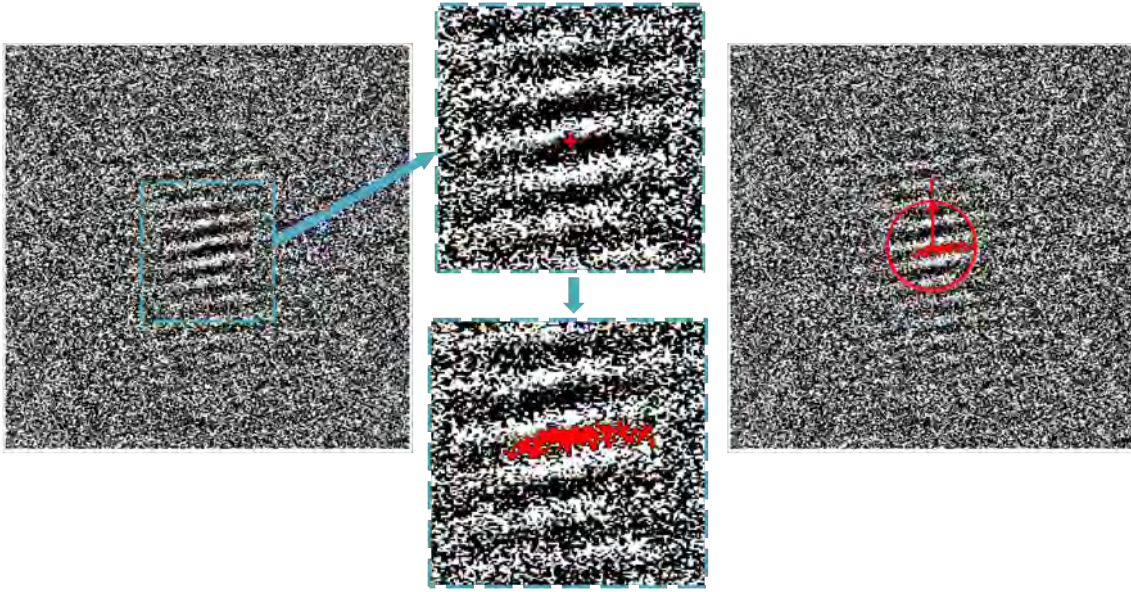
we do not have beforehand an idea of how the noise impacts the phase correlation matrix, we do have an idea of how it should look in the ideal noise-free case. If we recall the derivations of [Hoge2003] and [Balci and Foroosh2006], the discrete phase difference matrix (i.e.,  $\angle C(u, v)$ ) corresponds to a 2D sawtooth signal whose cycles determine the image shifts. Indeed, the phase difference matrix of a pure delta (see Figure 3.12) corresponds to a totally striped matrix according to the shifts in the spatial domain. As we have observed in the tests performed at the beginning of Section 3.2, the noise present in FLS images does not allow for a reliable estimation of the sawtooth cycles as they are typically corrupted by noise and cannot be accurately measured. The key idea here is that the area where the sawtooth pattern is perceivable (even if the stripes themselves are not "measurable") gives a good estimate of the frequencies that contribute to the correlation peak while the rest can be considered noise that will yield to spurious peaks once the phase correlation matrix is inverted back to the spatial domain. Therefore, assuming that we can somehow delimit the striped pattern area, we could place the cutoff frequency of the low-pass filter in its boundary in order to get a good trade-off between the minimization of noisy frequencies and the preservation of the peak.



**Figure 3.12:** (a) 2D Delta in the spatial domain centered at (10,10) (dilated for visualization purposes). (b) Its corresponding FFT.

In order to circumscribe the area of support of the correlation peak we take advantage of the fact the visible pattern will always emanate from the frequency origin. Hence we propose to use a simple but effective watershed type of segmentation [Meyer and Beucher1990] from the central point of the matrix  $\angle C(u, v)$ . The flood-fill segmentation spreads through the central stripe until the noise contaminates the pattern enough to prohibit its expansion. Thus, we will select the maximum coordinate of the segmented region as the radius to establish the filter cutoff.

In order to check that the described mechanism leads to the proper adaptation of the cutoff frequency, we have selected a hundred pairs of overlapping FLS images from different datasets and we have computed the phase correlation by applying the described frequency filtering. We have compared the correlation peak after the filtering with the automatically determined cutoff and the filtering using a range of all possible cutoff frequencies (expanding from 0 to 0.5 in 0.01

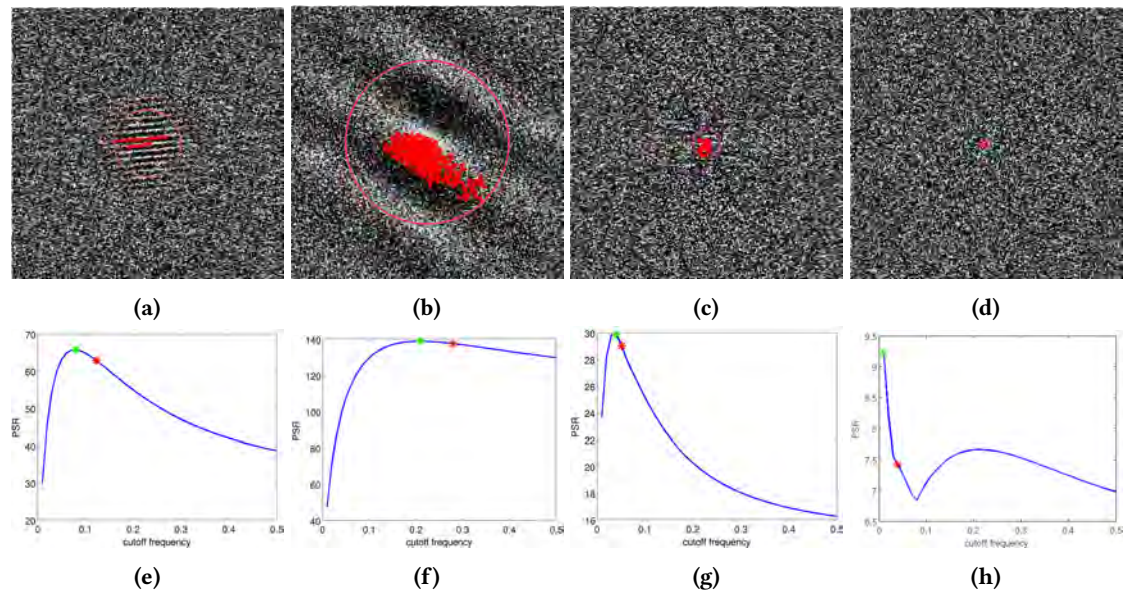


**Figure 3.13:** Adaptive determination of the filter cutoff frequency. A flood-fill segmentation from the origin of  $\angle C(u, v)$  establishes the filter cutoff radius.

intervals). For comparison purposes, we have defined a measure to quantify the peak sharpness through the Peak-to-Sidelobe Ratio (PSR):

$$PSR = \frac{peak - \mu}{\sigma} \quad (3.5)$$

which essentially measures how many standard deviations above the mean is the peak in the correlation output. Hence we define the optimal cutoff frequency as the one that leads to a higher PSR measure, indicating that we have reached the maximum compromise between suppressing the noisy frequencies and preserving the ones that contribute to determine the correlation. The average difference of the automatically selected cutoff with respect to the optimal one over the hundred pair-wise correlations is only of  $0.027 \pm 0.06$ . Figure 3.14 shows some examples of the obtained PSR curves for the different cutoff values, where it can be seen that the automatically selected cutoff falls very close to the maximum PSR. Moreover, we can see that the method leans toward the conservative side: it is always preferable to select a cutoff that does not filter the noise as optimally as possible (still being possible to detect the peak) rather than filtering more than necessary. Last column shows an example of two images that are actually not overlapping. In that case we can see that the selected cutoff is very small. However, since the images are totally different and there is no correlation (as indicated by the low PSR value), the actual peak result is irrelevant in this case.



**Figure 3.14:** Examples of automatically determined cutoff values (top row) for different pairs of overlapping images (a-c) and a non-overlapping pair (d). Second row evaluates the obtained peak sharpness once the corresponding phase correlation matrices have been filtered with a range of possible cutoff values. The green star shows the optimal cutoff leading to a highest PSR value and the red star shows the PSR value for the automatically detected cutoff.

### 3.2.2 Rotation estimation

Up to this point, we have covered the phase correlation approach for the estimation of linear shifts between FLS images. However, the recovery of rotation between frames must also be addressed. The inherent nature of sonar image formation suggests that mapping an area while maintaining the same orientation increases the chances of successful registration as intensity alterations due to viewpoint changes are minimized. In this sense, a lawn mower pattern where 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. The vehicle deploying the sonar device may not allow for the sway degree of freedom, or perhaps the area to cover does not conform to 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 ROVs as well, the pilot will most likely undertake a great number of rotational movements. Therefore, it is important to find a robust solution to estimate the rotation between pairs of FLS images so as to enable FLS mapping in more general situations and diverse environments.

Indeed, since most of the FLS registration techniques existing in the literature are feature-based (recall Sections 3.1.1 and 3.1.2), FLS rotation estimation has not received explicit treatment as it is intrinsically estimated from the feature matching step. Following the same reasoning introduced in Section 3.1.3 we want to explore global techniques that involve the use of all the image information without requiring the extraction of features, and at the same time, offer better



tolerance to noise and inhomogeneous intensity artifacts.

Hence, given two images  $f$  and  $g$  related by a roto-translation:

$$i_1(x, y) = i_2(x\cos\theta_0 + y\sin\theta_0 + t_x, -x\sin\theta_0 + y\cos\theta_0 + t_y), \quad (3.6)$$

where  $\theta_0$  and  $[t_x, t_y]$  are the relative rotation and translations respectively, we seek  $\theta_0$  so that we can correct the rotation between the images and afterwards solve for the translational shifts with the phase correlation introduced in the previous section.

It should be noted that in using this scheme, the accuracy of the rotation estimation step is crucial, as the translation estimation depends on the preceding rotation compensation.

In the following sections we describe four potential registration algorithms to solve the rotation estimation problem in a global manner.

### 3.2.2.1 The brute-force approach

According to the Fourier rotation property, a rotation of a function by  $\theta$  implies that its Fourier transform is also rotated by the same angle:

$$I_1(u, v) = I_2(ucos\theta_0 + vsin\theta_0 + x_0, -usin\theta_0 + vcos\theta_0 + y_0)exp^{-j(ut_x+vt_y)}. \quad (3.7)$$

Then, the rotation angle between  $i_1$  and  $i_2$  can be computed as the angle  $\theta$  for which  $I_1$  and  $I_2$  yield to the maximum correlation. However this implies the computation of an image rotation, its Fourier transform, and the subsequent correlation for every angle hypothesis, becoming computationally expensive.

Similar approaches have been explored in the domain of the Radon transform [Li et al.2007, Costello2008], where the dimensionality of the rotation estimation problem is reduced to 1D projections. In order to estimate a rotation between two images, it is required to compute 1D correlations between the reference Radon transform vector and all the different vectors corresponding to each of the potential rotation angles. Again, this would require to evaluate a range of angle hypothesis covering all the sonar's FOV, generating a hardly satisfiable trade-off between accuracy and computation time.

Hence, although these kind of approaches could certainly be valid to estimate the rotation between two FLS frames, they are opposed to our objective of finding an efficient FLS registration method and they are not further considered.

### 3.2.2.2 Fourier-Mellin transform

One of the most popular global registration 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 et al.1994, Reddy and Chatterji1996]. This method is indeed an extension of the phase correlation to estimate rotation and scale changes by using the log-polar domain. In our case we will disregard the log factor since, as stated earlier, there is not scale ambiguity between FLS images.

The Fourier transform of Equation (3.7) can be expressed in polar coordinates as:

$$I_1(r, \theta) = I_2(r, \theta + \theta_0) e^{-i(ut_x + vt_y)}. \quad (3.8)$$

As can be observed in Equation (3.8), the rotation in the spatial domain is mapped as a linear shift in the angular direction of the polar domain. According to the Fourier shift property, we can see also that the translational displacements affect only the phase spectrum, so the rotation can be determined independently using the magnitude of the polar 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 phase correlation as described in Section 3.2.1, and leads to two possible solutions ( $\theta$  and  $\theta + \pi$ ) that can be disambiguated by solving in both cases for the subsequent translation and keeping the rotation that leads to the highest correlation peak.

As introduced in Section 3.2.1.1 it is necessary to mask the images to avoid edge effects caused by the sonar fan-shaped footprint, as well as to window, in this case, the polar magnitudes of the Fourier transforms. Under this approach, phase correlation is not applied on the original images, but on the polar magnitude of the Fourier transforms which are images characterized by a low structural nature (particularly in the case of sonar modality). Moreover, these images are suffering from inaccuracies introduced by the conversion to the polar domain as the data from the Cartesian grid needs to be sampled to obtain a regular  $r - \theta$  grid. For low frequencies, which generally contain most of the energy, the sampling is based on very few values and thus can introduce significant approximation errors.

Given that Fourier-Mellin transform is considered a reference technique for global rotation estimation and is widely popular on optical images [Bülow et al.2009, Schwertfeger et al.2010], its performance on FLS images will be evaluated in Section 3.3.1.

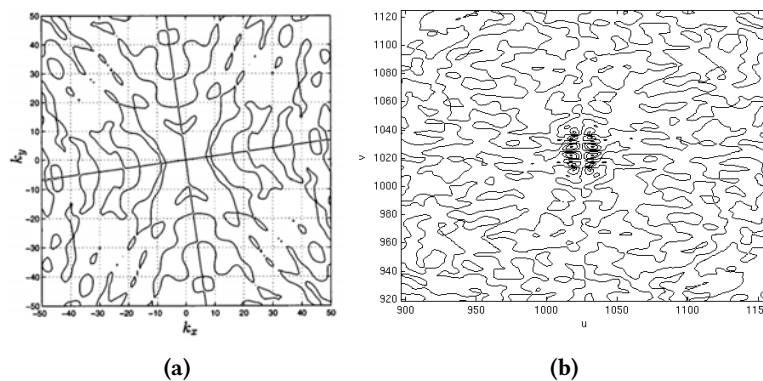
### 3.2.2.3 Angular Difference Estimation

Lucchese and Cortelazzo introduced another Fourier-based method for rotation estimation [Lucchese and Cortelazzo2000]. Their approach is not based on a correlation strategy but on a specific

property that enables the estimation by working directly in the Cartesian domain, bypassing the need for any polar conversion. Given two images, they have shown that the difference between the magnitudes of the Fourier transforms of the first image and the mirrored version of the second has a pair of orthogonal zero-crossing lines. The orientation of these lines with respect to the frequency axis is shown to correspond to half the rotational angle [Lucchese and Cortelazzo2000]. The authors acknowledge that the method works better for images that are fully overlapping, although they show that the orthogonal lines can still be recovered (with decreased accuracy) when using images with less overlap such as in the presence of translations combined with rotations.

Several techniques have been proposed for the detection of this cross-pattern including the use of a multiscale Hough transform [Lucchese and Cortelazzo2000] or the detection through the definition of an Angular Difference Function [Keller et al.2005].

Figure 3.15 shows an example from [Lucchese and Cortelazzo2000] with two input images (flipped accordingly) and its corresponding Fourier transform magnitude difference. The cross pattern is distinctive and thus the rotation angle can be easily computed. Figure 3.15b shows the same operation on two FLS images (previously masked to avoid edge effects). The contour of the difference image around the center reveals an intricate pattern where the cross pattern cannot be appreciated, even though the input images have a high overlap. The noise and the weak structure of FLS images dilute the presence of the orthogonal crossing and therefore rotation estimation through these types of techniques do not seem a valid option for FLS imagery.



**Figure 3.15:** Contour plot of the angular difference between the Fourier magnitude of one image and the mirrored version of the other. Example extracted from [Lucchese and Cortelazzo2000] (b) Example on two FLS images.

### 3.2.2.4 Direct Polar Estimation

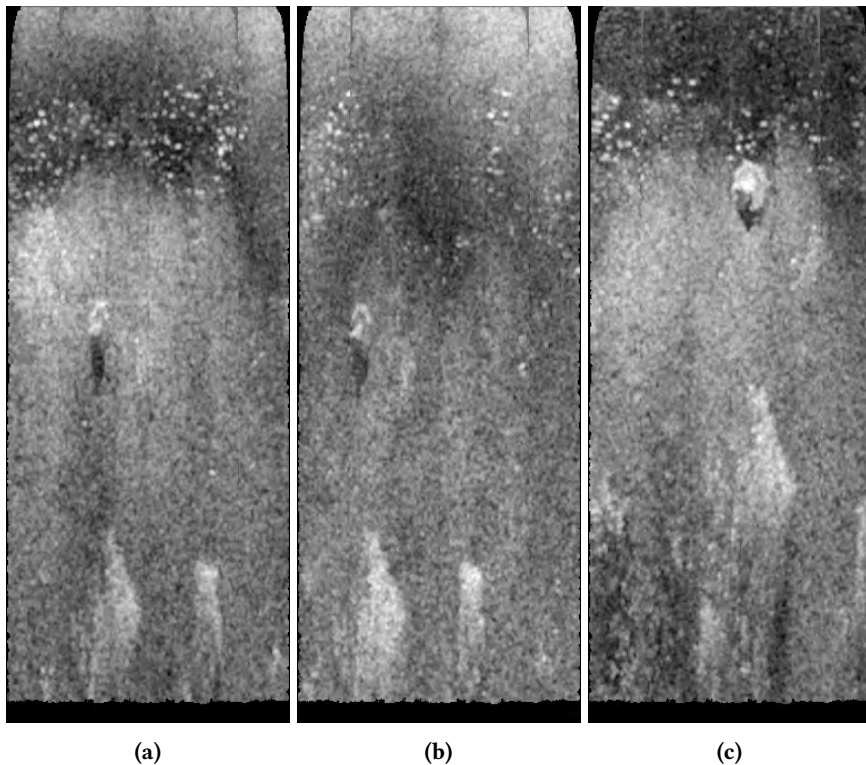
The previously described Fourier-Mellin approach disassociates rotation from translation and converts it to a shift by utilizing the polar magnitude of the Fourier spectrum. In general, the Cartesian-to-polar coordinate transformation of the Fourier spectrum is a numerically sensitive

operation which is subject to approximation errors. In view of that, approaches have appeared [Averbuch et al.2006] to directly compute the Discrete Fourier Transform (DFT) on a pseudo-polar grid without requiring interpolation on the Fourier magnitude spectra. However, regardless of how the polar conversion is computed, this still implies performing phase correlation on the Fourier magnitude data. While video images typically present dominant edges and accentuated features that give rise to strong components in the magnitude spectra, FLS images exhibit, in general, a low structural spectral magnitude. This lack of distinctive traits combined with the fact that the Fourier magnitude is, by itself, a representation of oscillatory nature, does not lead to solid foundation on which to perform phase-correlation.

Considering all this, we examine the option of applying phase correlation directly on the spatial domain data, estimating the rotation as a shift displacement in the across-range direction of the polar images. 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 the magnitude of the Fourier transform. Two important considerations must be highlighted here. First, rotation about the sonar origin becomes translation along the angle axis in the polar images delivered by the sonar. However, if the center of rotation is not located on the sonar coordinate frame  $\{S\}$  (e.g., in the center of the vehicle), the polar image must be recomputed so as the origin of its axes takes into account the 2D transformation that relates  $\{S\}$  with respect to the center of rotation. Second, regardless of the case of pure rotations, this scheme allows only to recover the rotation in an approximate manner: when working directly 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, 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 introducing significant errors under this scheme. Moreover, there are many cases where rotations are not combined with translations (the vehicle stops, rotates and then continues) yielding a pure translation in the polar domain. The major drawback arises when attempting to match temporally distant frames that present significant shifts such as in loop-closing situations. In these cases, the proposed strategy for rotation estimation is prone to introduce inaccuracies in the estimated angle. This, in turn, has an impact in the number of encountered loop closures, being the loop closures that involve more overlap and smaller orientation changes more likely to be successfully registered.

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  $[-\frac{FoV}{2} : \frac{FoV}{2}]$  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 FOV. 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  $[-\frac{FoV}{2} + 180 : 180 + \frac{FoV}{2}]$  degrees.



**Figure 3.16:** Distortions on the estimation of the rotation angle as a shift in the polar images. (a) Reference frame. (b) Frame undergoing a pure rotation, observe that regardless of the intensity artifacts, the rotation is mapped to a shift in the horizontal direction. (c). Frame undergoing both a rotation and a large translation. Note that although it is still possible to compute the displacement in the horizontal direction, the translation has induced more changes on the image content and therefore the estimation of the rotation will certainly be less accurate.

In Section 3.3.1, the performance of this approach will be quantitatively compared to the traditional Fourier Mellin technique that estimates rotation by using the polar Fourier transform magnitudes of the images. We will see that despite the mentioned limitations, it outperforms the Fourier Mellin methodology when applied on FLS images, and therefore is the rotation estimation

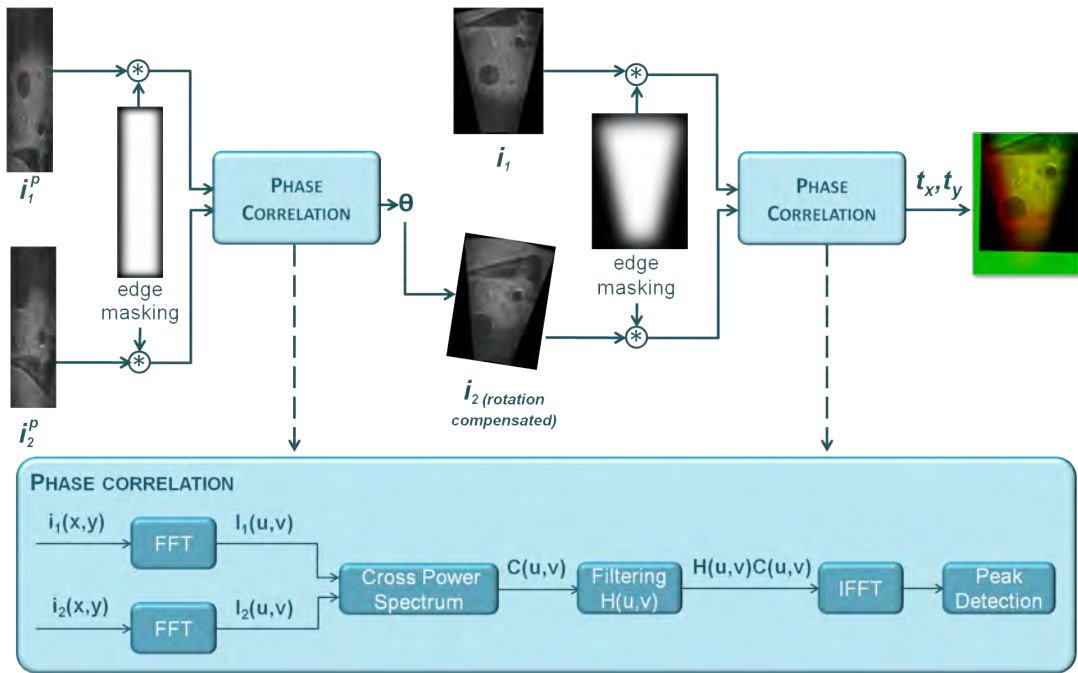


Figure 3.17: Overall registration pipeline.

mechanism employed in our registration pipeline.

### 3.2.3 FLS Registration Outline

After describing the phase correlation approach to estimate the translation and rotation for FLS images, we summarize here the full process to register two frames. The flowchart of Figure 3.17 outlines the proposed registration method. The sonar frames in polar coordinates ( $i_1^p, i_2^p$ ) are first masked by a cosine window to avoid edge-effects arising from the image's boundaries. Using as input the transforms of these images, phase correlation is applied following the process described in Figure 3.9c. The shift detected in the  $x$  direction provides an estimate of the rotation angle  $\theta$  between the images in Cartesian coordinates. Afterwards,  $i_2$  is rotated to compensate for the detected angle and after masking both  $i_1$  and the rotated  $i_2$  with the corresponding Cartesian mask, phase correlation is applied again to finally obtain the translations in the  $x$  and  $y$  directions that align the two images.

## 3.3 Quantitative Comparisons

In this section, we provide quantitative comparisons to prove the performance of the proposed registration method. Two aspects are analyzed: first, the performance of the proposed rotation estimation method against the traditional Fourier-Mellin transform, and second, the performance of the proposed Fourier-based registration method compared against other state-of-the-art FLS



registration techniques, including a feature-based method at pixel level and one at region level.

In order to carry out the comparisons, 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. An example frame of each dataset can be seen in Figures 3.18a-3.18c. The first dataset is comprised of 944 sonar frames gathered with an ARIS sonar [Sound Metrics ARIS2013] inside a harbor. The FLS was mounted on a pole together with 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, range resolution is 8 mm/pixel and angular resolution 0.2 deg. The second dataset consists of 1176 sonar frames gathered with a DIDSON [Sound Metrics DIDSON2013] in a dock environment. The sonar performed a 360 deg scan with steps of 0.3 deg 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 Vehicle (ASV) performing both rotational and translational motions. Similar to the first dataset, the FLS was pole-mounted together with a GPS unit. Range and angular resolution are 6 cm/pixel and 0.3 deg respectively. Therefore, the estimated translations and rotations will be compared using as ground truth the GPS positions in the first and third dataset and the fixed mechanical tripod step in the second dataset. It is worth noting that the employed GPS units are high-precision Real-Time Kinematic (RTK) GPS that also deliver an accurate heading by employing a setup with two antennas. Moreover, the reported errors are the average of large number of registration results. In this way, we consider that the effect of any possible GPS errors over the reported mean errors is negligible.

### 3.3.1 Comparison of the simplified rotation estimation method versus the Fourier-Mellin Transform

We have used the described datasets to compare the proposed rotation estimation method (Section 3.2.2.4) with the traditional Fourier Mellin approach (Section 3.2.2.2). For each dataset, two different tests have been performed. The first one 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 distant 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.

Table 3.1 presents the mean and maximum rotation errors with respect to the ground truth when estimating the rotation between consecutive frames while Table 3.2 shows the results for

the case of distant images.

	Fourier-Mellin		Directly on Polar Images	
	<i>Mean error(deg)</i>	<i>Max error(deg)</i>	<i>Mean error(deg)</i>	<i>Max error(deg)</i>
<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 3.1:** Comparison experiments between rotation estimation methods when registering consecutive frames.

	Fourier-Mellin		Directly on Polar Images	
	<i>Mean error(deg)</i>	<i>Max error(deg)</i>	<i>Mean error(deg)</i>	<i>Max error(deg)</i>
<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

**Table 3.2:** Comparison experiments between rotation estimation methods when registering distant frames.

In all cases, even when estimating the rotation of distant frames, the estimation through direct phase correlation on the polar images leads to lower average errors than performing the estimation on the polar magnitude of the image's Fourier transforms. 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. This testifies to the fact that the noise and the low structural nature of sonar images impede the robust correlation of the polar spectrum magnitudes carried out in the Fourier-Mellin approach. Therefore, as argued in Section 3.2.2.4 the use of the simplified rotation estimation method is better in most of the cases.

### 3.3.2 Comparison against feature-based registration

The performance of the proposed Fourier-based method (as sketched in Figure 3.17) is compared against two other state-of-the-art FLS registration techniques: one feature-based method at pixel level and one feature-based method at region level. Regarding the former, we have implemented an approach similar to the one of [Kim et al.2005] and [Negahdaripour et al.2005] where Harris corners are used as a feature detector and the matching is performed through correlation of local sub-patches around the features. Finally, the transformation estimation is performed after a RANSAC process to reject outliers. Although there are many modern feature detectors that have been developed to work under severe viewpoint changes and estimate more complex transformations (e.g., SIFT [Lowe2004] Speeded-Up Robust Features (SURF) [Bay et al.2006]) it is recommended to use the lowest degree of invariance required for the application at hand [Tuytelaars and Mikolajczyk2008]. Therefore, the selection of Harris detector is appropriate in this sense as it is only translation and rotation invariant. It is important to note that although several researchers



have pointed out that feature-based registration at pixel level does not perform well on FLS images, we make the effort to include it in this comparison as to the best of our knowledge there is no quantitative evidence of that claim in the literature.

Regarding the feature-based approach at region level, we have selected, from the two existing methods in the literature [Johannsson et al.2010, Aykin and Negahdaripour2012], 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.

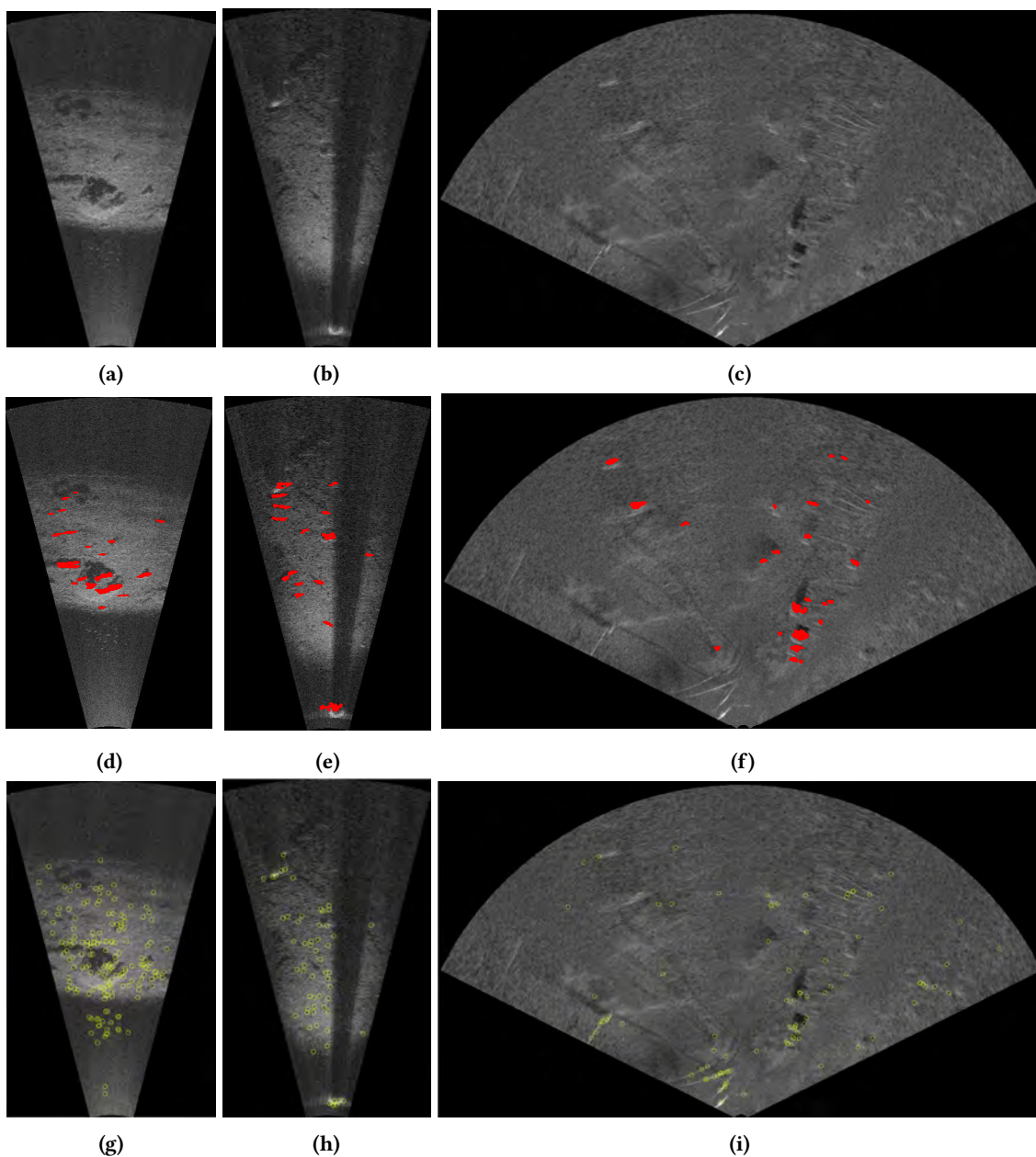
We have implemented the technique of Johannsson *et al.* following their steps as described in [Hover et al.2012]. The feature extraction process applies median smoothing on the image followed by gradient computation. The gradient is computed as the difference between a value and the mean of its  $n$  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 4 overlapping grids shifted half a cell. The NDT implementation of the Point Cloud Library (PCL) [Point Cloud Library2013] has been used for this step. Following the same procedure as the authors, the NDT optimization is performed several times with different initialization points.

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 threshold 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.

Figure 3.18 shows three images illustrative of each dataset together with examples of extracted features. Again, for each dataset two different tests have been performed, one attempting the registration of consecutive frames and one with more distant frames.

	Mean error			Max error			% of non-computed pairs	Repeatability rate
	$t_x(m)$	$t_y(m)$	$\theta(deg)$	$t_x(m)$	$t_y(m)$	$\theta(deg)$		
<b>Dataset 1</b>	3.9	3.6	2.8	14.5	12.5	18.9	43	0.71
<b>Dataset 2</b>	3.6	2.2	3.2	17.1	15.6	23.6	18	0.79
<b>Dataset 3</b>	4.5	3.8	4.5	15.8	16.8	14.1	28	0.71

**Table 3.3:** Results of the feature-based registration at pixel level when registering consecutive frames.



**Figure 3.18:** Example of dataset frames used in the quantitative comparisons. (a)-(c) Original frames. (d)-(f) Example of extracted features with Johannsson *et al.* method. (g)-(i) Example of extracted features with Harris detector.

	Mean error			Max error			% of non-computed pairs	Repeatability rate
	$t_x(m)$	$t_y(m)$	$\theta(deg)$	$t_x(m)$	$t_y(m)$	$\theta(deg)$		
<b>Dataset 1</b>	4.2	3.9	5.8	16.1	18.5	14.3	74	0.56
<b>Dataset 2</b>	3.6	3.7	4.5	14.9	13.6	20.4	80	0.58
<b>Dataset 3</b>	6.6	4.9	7.3	22.2	38.8	41.1	75	0.48

**Table 3.4:** Results of the feature-based registration at pixel level when registering distant frames.

Tables 3.3 and 3.4 summarize the results of both tests for the feature-based registration at pixel level, showing the mean and maximum errors for the rotation and translations in each dataset. In a high percentage of cases it has been impossible to compute the transformation relating the two frames due to an insufficient number of inlier matches, especially in the case of distant frames. Therefore the error results in tables 3.3 and 3.4 refer only to the cases where the transformation could be computed. In view of this poor number of inlier matchings, an interesting parameter to evaluate is the repeatability of the feature detector on FLS images. Repeatability is an important characteristic of a feature detector, which indicates the ability to detect the same feature in two different images (e.g., if the scene is revisited from a different viewpoint or under different illumination conditions). In order to perform this evaluation, we have computed the repeatability rate following the measure described in [Schmid et al.2000]. By knowing the homography that relates one image to the other according to the ground truth, the detected features of the common image area are searched in a small neighborhood to determine whether they are repeated features or not. In our tests we have set this neighborhood to 5 pixels, which ensures that the probability of two points falling accidentally within this neighborhood is very low. 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.

The repeatability rates for each dataset are presented in the last columns of tables 3.3 and 3.4. The rates are relatively low compared to usual values reported in video images for the same Harris detector (which are typically above 90% [Schmid et al.2000]). It can also be observed that the repeatability decreases in the case of distant frames. This suggests that the inhomogeneous resolution, the low signal to noise ratio and the intensity variations due to viewpoint changes complicate the detection of the same corner points in two different views. Moreover, in some occasions even though the same feature is correctly located in both images, they are not matched together as there exist other features in similar image areas that eventually have a higher correlation score, leading to incorrect matches.

This combination of low repeatability and low distinctiveness generates a high number of outliers. Although RANSAC does a good job in rejecting them, that leads often to a situation where very few inliers persist, not being enough to estimate the underlying transformation. In other cases, not all outliers can be filtered thus producing erroneous estimations which yield the high errors shown in tables 3.3 and 3.4.

Tables 3.5 and 3.6 summarize the results of the two tests for the registration at region level and the proposed Fourier-based registration.

	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>Dataset1</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>Dataset2</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>Dataset3</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

**Table 3.5:** Comparison experiments between the Fourier-based and region-based registration methods when registering consecutive frames.

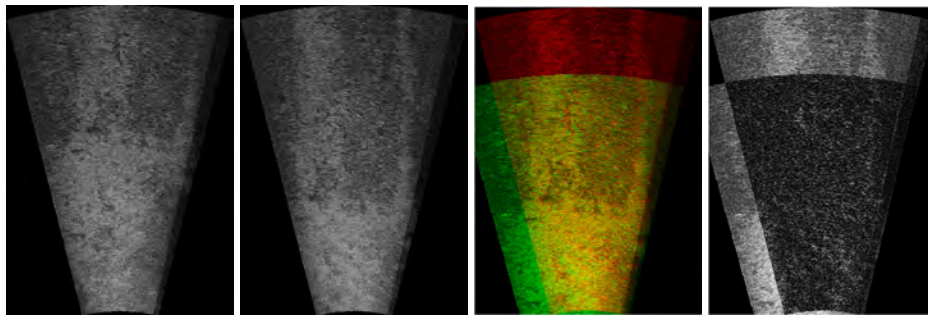
	Region-based						Fourier-based					
	Mean error			Max error			Mean error			Max error		
	$t_x(m)$	$t_y(m)$	$\theta(deg)$	$t_x(m)$	$t_y(m)$	$\theta(deg)$	$t_x(m)$	$t_y(m)$	$\theta(deg)$	$t_x(m)$	$t_y(m)$	$\theta(deg)$
<b>Dataset 1</b>	0.60	0.26	1.34	5.62	3.63	12.5	0.35	0.24	1.15	1.05	1.11	3.91
<b>Dataset 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

**Table 3.6:** Comparison experiments between the Fourier-based and region-based registration methods when registering distant frames.

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 of 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 features in this dataset -sparse and weak- are 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 higher error rates. In particular, the errors for the region-based method have especially increased with respect to their counterparts in the first experiment. In the first two datasets, the smaller overlap and the narrow aperture of the sonar in the azimuth direction ( $\sim 30^\circ$ ) causes significant features to eventually drop out of the field of view, leading to an insufficient number of common features to perform the NDT alignment in a reliable manner. Moreover, 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. 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.

Regarding the third dataset, the errors have also increased under both registration methodologies compared to the previous test. For the region-based method, the error of the rotation estimation, is especially high, which also has an impact in the translation errors. Indeed, when comparing distant frames rotation is the motion most affected by intensity alterations. Note that



**Figure 3.19:** (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 colour 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.

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 fail to 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 Johannsson *et al.* method testifies to its better performance in these situations.

As all the analyzed sequences presented feature-rich environments, a different example is introduced to highlight the difficulties of region-based techniques in environments with a scarcity of features. Figure 3.19 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 consecutive and non-consecutive 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 inside a mosaicing framework.

### 3.3.3 A note on computational cost

The performance of the methods with regards to the computational cost is also worth noting.

The most demanding operation of the proposed registration method is the computation of

the DFTs. This computation, implemented through the FFT algorithm, requires  $O(MN \log_2 MN)$  operations for each 2D transform, where  $M$  and  $N$  are the number of pixels in each image axis. The current implementation of the registration algorithm, coded in Python and making use of the ANFFT libraries [ANFFT Package2013], consumes approximately 60 ms per pairwise registration in an Intel i7 at 3.4 Mhz, considering typical image sizes under 1024x1024 pixels and a single-threaded execution. Under the same machine, the feature-based registration using Harris corners takes an average of per 150 ms per registration. Finally, our implementation of the Johannsson *et al.* technique takes approximately 6 times longer than the Fourier-based registration, the major part of the time being consumed by the NDT optimization process.

A further advantage of the Fourier-based registration is that given an image size the computation time is constant, while in feature-based methods computation time fluctuates depending on the number of found features.



# 4

## GLOBAL ALIGNMENT

---

The registration method described so far is intended to compute the relative transformation between pairs of overlapping FLS 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 relative transformations over long sequences is prone to cumulative error [Smith and Cheeseman1986]. Therefore, we apply a global alignment technique to obtain a consistent set of absolute transformations that will enable the subsequent mapping of the images in a composite mosaic. We start the chapter by providing a brief overview of related work in global alignment, from which we determine that the most suitable approach is to face our problem as a pose-based graph optimization. We then present the general formulation of the pose graph and we give insights of the type of constraints that might be incorporated. Next, we describe two key aspects of the graph optimization front-end: first, the uncertainty estimation of the registration constraints and second, the methodology to select which registrations should be attempted and which links are to be included in the graph for both offline and real-time processing.



## 4.1 Related work

Global alignment techniques have been extensively studied to constrain the cumulative registration error in mosaicing applications [Elibol2011]. Global alignment usually requires the minimisation of an error term, which is defined from image correspondences between overlapping image pairs. Bundle Adjustment (BA) [Triggs et al.2000] is arguably the most popular method to perform global alignment for both photomosaicing and 3D reconstruction applications. BA methods simultaneously optimize a set of camera poses and feature points using non-linear least square methods to find the optimal camera motion parameters from which absolute homographies can be computed. A similar idea is applied in [Kim et al.2006], which to the best of the author's knowledge is the only work applying the notion of global alignment in a FLS mosaicing example. In there, the authors minimize the registration error of subsequences of the main image stream by registering all its frames to a selected anchor frame in the subsequence. The alignment of these sequence portions is optimized by minimizing the reprojection of feature points between the images. However, note that this process can not be fully considered global alignment as only ensures local consistency between the anchored neighborhood and registration error can still be accumulated between subsequences.

Bundle adjustment techniques from the computer vision community can be seen as an equivalent to the smoothing and mapping approaches employed in solving landmark-based SLAM: they are both based in the joint estimation of poses (either from a vehicle or from a camera) together with landmarks/visual image features. Indeed, globally aligning a set of image poses in the mosaic plane tightly connects to the classical SLAM problem where the pose of the vehicle along a trajectory is to be estimated according to measures on the environment.

During the past decades, filtering approaches, including Kalman filter and its variants [Smith and Singh2006], have been proved very successful in that purpose. However they are designed to track the current pose at a given time and can not provide an accurate estimate of the full trajectory, which is an essential requirement for mosaicing purposes. Actually, the strength of the global alignment arises from re-visiting an already mapped area, in the so-called loop-closure situations. Even that delayed information can be also handled by some implementations [Zhang and Bar-Shalom2011, Ranganathan et al.2007], standard filtering solutions cannot back-propagate the information of loop closure constraints to previous pose estimates as these loop closures may impact states far back in time that are no longer represented.

Instead, smoothing approaches keep past states in the estimation, solving the so-called full SLAM problem or view-based SLAM. These algorithms were introduced back in the 90's by Lu and Milios [Lu and Milios1997], and have been further explored by many researchers [Dellaert and Kaess2006, Eustice et al.2006, Kaess et al.2008]. It has been shown that although the state space is extended by incorporating the trajectory, the structure of the problem remains sparse and can be solved very efficiently. We refer the reader to [Dellaert and Kaess2006] for an excellent

explanation about the problem sparsity and how the structure of SLAM problems in general lead to efficient smoothing solutions. One relevant example to this thesis is the work of Walter *et al.* [Walter et al.2008], in which a feature-based SLAM is implemented based upon features manually selected from FLS images. The authors use an Exactly Sparse Extended Information Filter (ESEIF) in order to keep track of the features and the robot poses along all the trajectory.

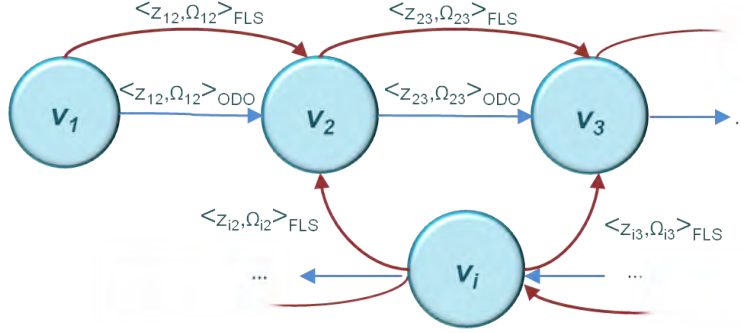
A method that has recently gained a lot of popularity are the so-called pose-based graphs, which offer a natural representation for smoothing problems. Pose graphs are abstract representations of an optimization problem consisting of pose nodes, which represent the variables to be estimated, and edges, that represent constraints on the variables they connect (Figure 4.6). Essentially, the graph nodes (or vertices) encode the trajectory map while the edges encode measurements constraints. Therefore the explicit estimation of landmark locations is omitted by integrating relative constraints between pairs of poses. The pose graph structure has usually a sparse connectivity and its optimization can be efficiently implemented. Indeed their increase in popularity is partly thanks to the new availability of efficient algorithms to solve the underlying graph optimization problems.

Pose graphs have been shown to produce consistent maps due to their ability to correct prior navigation error and re-linearize around the correct trajectory opposite to filtering approaches that are prone to introduce linearization errors. Successful examples of pose-based graphs in the underwater domain include photomosaicing [Beall et al.2011], combination of visual and multibeam sonar measurements [Kunz and Singh2013], 3D sonar reconstructions of complex environments [VanMiddlesworth2014] and, more compelling to the work presented in this thesis, SLAM using FLS image registration constraints [Hover et al.2012].

Altogether, a pose-based graph formulation seems the more suitable approach for performing global alignment in our FLS mosaicing pipeline. First, because of the ability of estimating the full trajectory and handle loop closures which are essential to achieve global alignment and map all images into a consistent mosaic. Second, because of the suitability of the representation. Since we use an area-based registration method we can not use the feature reprojection error for optimizing the image alignment and therefore an optimization that omits explicitly the environment features becomes convenient. And finally, because of their efficiency when dealing with a large number of poses (in our case image poses) which is the key to achieve better results in terms of coverage and SNR of the final mosaic.

## 4.2 Pose-based Graph Definition

We define a graph whose vertices represent the position of observed sonar images and whose edges are pose constraints obtained either from pairwise image registrations or from vehicle odometry measurements (if navigation sensors are available).



**Figure 4.1:** Depiction of the pose-based graph structure. In our case, nodes or vertices are image poses. Red edges represent constraints derived from the pairwise registration between FLS frames and blue edges are navigation constraints, that may or may not be available.

We consider two different situations throughout the global alignment procedure presented here: working exclusively with FLS imagery or also being able to incorporate odometry measurements from navigation sensors. The high frame rate -and therefore the large overlap- of FLS imagery allows us to contemplate the case of dealing only with image constraints. Under common refresh rates and vehicle speeds, multiple consecutive and non-consecutive constraints can be potentially found. Eventually, there are enough to link all images and ensure consistency in the path without requiring other type of edges. This extends the applicability of the mosaicing pipeline to situations where the sonar is deployed from vehicles with reduced sensor suites, which is usually the case in ROVs operations. Besides, there are other situations where acquiring correct navigation data might be difficult (e.g. using a compass close to magnetic disturbances) and being able to generate a mosaic only out of FLS data can be of interest.

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 is available, the vertices can be initialized using the pose estimates from the vehicle dead-reckoning information.

Let  $\mathbf{z}_{i,j}$  and  $\Omega_{i,j}^z$  be the mean and information matrix respectively of the spatial transformation from image  $i$  to image  $j$  obtained from:

- applying the registration algorithm on the image pair  $(i, j)$   
or
- the vehicle odometry increments between the acquisition times of image  $i$  and  $j$ .

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:

$$\mathbf{E}(\mathbf{v}_i, \mathbf{v}_j, \mathbf{z}_{i,j}) = \mathbf{z}_{i,j} \ominus \hat{\mathbf{z}}_{i,j}(\mathbf{v}_i, \mathbf{v}_j) \quad (4.1)$$

where  $\ominus$  is the inverse of the usual motion composition operator for rigid body motions in the 2D Euclidean space [Smith et al.1987].

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  $\mathcal{C}$ :

$$F(\mathbf{v}) = \sum_{(i,j) \in \mathcal{C}} \mathbf{E}(\mathbf{v}_i, \mathbf{v}_j, \mathbf{z}_{i,j})^T \boldsymbol{\Omega}_{i,j}^z \mathbf{E}(\mathbf{v}_i, \mathbf{v}_j, \mathbf{z}_{i,j}), \quad (4.2)$$

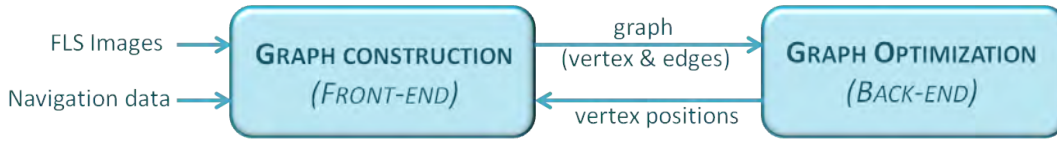
$$\mathbf{v}^* = \arg \min_{\mathbf{v}} (F(\mathbf{v})) \quad (4.3)$$

If a good initial guess of the parameters is known, a numerical solution of Equation 4.3 can be obtained using a non-linear least squares solver such as the popular Levenberg-Marquardt algorithm [Moré1978]. The first order Taylor expansion around the current guess  $v^*$  is used iteratively to approximate Equation 4.3 and optimize the local  $v$  increments by solving the resulting sparse linear system. The starting point for the next iteration is obtained by adding the optimized increments to the initial guess using the motion composition operators and the process is repeated until the increments are smaller than a predefined threshold a predefined number of iterations has been completed.

In the last years, the SLAM community has begun to provide many graph back-ends that contain optimizers to solve the nonlinear optimization problem expressed by pose graphs: TORO [Grisetti et al.2009],  $\sqrt{SAM}$  [Dellaert and Kaess2006], iSAM [Kaess et al.2008], Sparse Pose Adjustment [Konolige et al.2010], iSAM2 [Kaess et al.2011], SLAM++ [Polok et al.2013] are some of the examples. In our implementation the minimization of Equation 4.3 is solved using the General Framework for Graph Optimization (G2O) back-end [Kummerle et al.2011], which is designed to provide an easily extensible yet efficient implementation. While there is a large body of ongoing research devoted to provide more efficient and robust graph back-ends, it is not the focus of this thesis to contribute in this area. Instead, we make use of the selected back-end as a tool and we direct our efforts in providing a reliable graph front-end (see Figure 4.2). The graph front-end is the responsible to construct a topologically correct graph from the available sensor information, in our case the FLS images and, if present, the vehicle navigation data. If that fails, for instance due to the introduction of erroneous loop closure constraints, the subsequent optimization is likely to produce deficient solutions.

Thus, in order to provide a well-defined graph to the optimization back-end there are two key tasks to be addressed by the the front-end:

- Compute an information matrix ( $\boldsymbol{\Omega}^z$ ) for each measurement so that it will appropriately weight the contribution of the edge constraint in the overall error function.



**Figure 4.2:** Interaction between graph front-end and back-end.

- Decide which constraints must be included in the graph. This requires the prior selection of candidates to explore data association, i.e., the candidate frames on which registration will be attempted.

Next sections will provide insights in each of these tasks.

### 4.3 Estimation of the Constraints Uncertainty

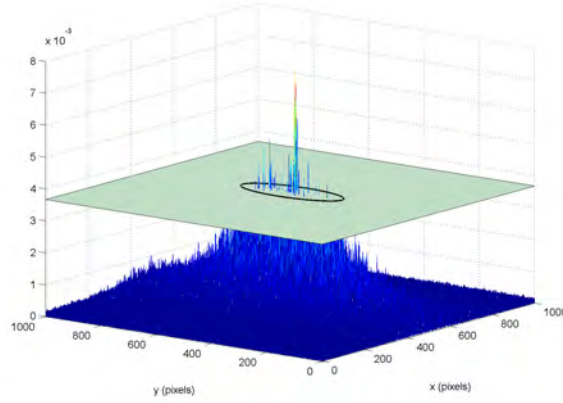
The described pose-graph formulation requires establishing an information matrix  $\Omega^z$  for every edge constraint in the graph. In the following, we show how this matrix is computed for both registration and odometry constraints.

#### 4.3.1 Registration Constraints

In order to quantify the degree of confidence in the pairwise alignment we derive a heuristic from the proposed Fourier-based registration method. Recalling the description of the method in Section 3.2, the inverse Fourier transform of the phase correlation matrix can be used to measure the degree of congruence between two images. The key insight is that the amplitude and extent of values surrounding the main peak of the correlation surface account for localization inaccuracies in the registration.

Our approach consists in thresholding the phase correlation surface at a given amplitude and compute the standard deviations of the  $x$  and  $y$  coordinates of the matrix cells that exceed the threshold, as depicted in Figure 4.3. The threshold is set to half the power of the main peak, so that only significant values are taken into account.

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:



**Figure 4.3:** Representation of the proposed heuristic to compute the uncertainty of the registration from the phase correlation matrix values.

$$\mathbf{\Omega}^z = \begin{bmatrix} (\sigma_x \delta_r)^2 & 0 & 0 \\ 0 & (\sigma_y \delta_r)^2 & 0 \\ 0 & 0 & (\sigma_r \delta_\theta)^2 \end{bmatrix}^{-1} \quad (4.4)$$

A similar heuristic was proposed in earlier work by Pfingsthorn *et al.* [Pfingsthorn et al.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 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 only those values that are above half power of the main peak (i.e., values within 3dB 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 available ground truth to evaluate how the different uncertainties adapt to the real registration results. Hence, if a measure explains well the uncertainty of the data, the ground truth registration result will be within the  $\sigma$  bounds of the determined covariance matrix with a given confidence interval. We have selected 4000 frames from a dataset with available GPS ground truth and we have computed the pairwise registration of each frame against 20 other neighbor frames with our proposed Fourier-based method. This amounts a total of 80000 pairwise registrations, involving cases of high overlap where the resulting peak surface has a pronounced peak and registrations with more distant frames which lead to more weak peaks. For each of this registrations, four uncertainty estimates have been computed on the correlation matrix: our proposed heuristic and the heuristic of Pfingsthorn *et al.* 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  $1024 \times 1024$ . In this way, we have chosen small  $K = 30$ , medium  $K = 60$  and large  $K = 120$  values.

The first parameter to evaluate is the percentage of times the true registration result falls within the  $\sigma$  bounds of the different uncertainty measures, computed at a confidence interval of 99%. The second, is the average area of the computed uncertainty ellipses. As shown in Table 4.1, the uncertainty computed with the metric of Pfingsthorn *et al.* is clearly dependant on the window size parameter  $K$ .

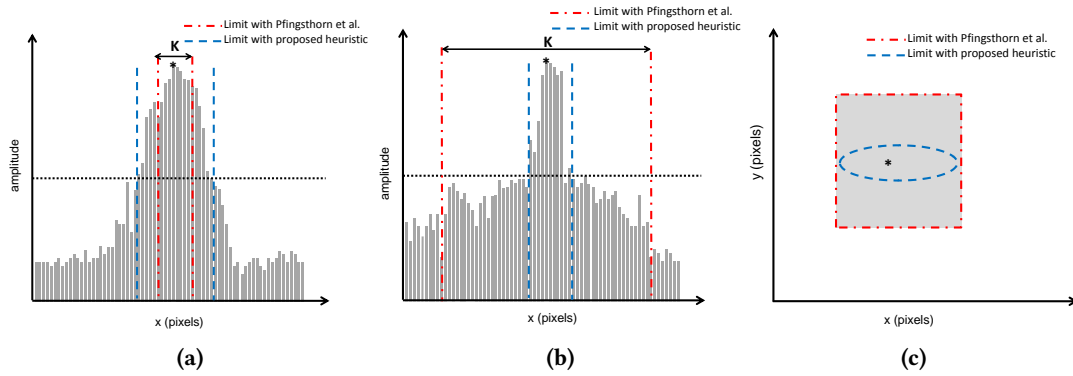
Method	% within $3\sigma$ bounds	Average ellipse area( $m^2$ )
<b>Pfingsthorn, K=30</b>	63%	2.42
<b>Pfingsthorn, K=60</b>	89%	26.08
<b>Pfingsthorn, K=120</b>	95%	195.08
<b>Proposed</b>	88%	5.86

**Table 4.1:** Evaluation of uncertainty measures. Percentage of true registration results comprised inside  $3\sigma$  bounds of the different uncertainty estimates and the corresponding average of the uncertainty ellipse areas.

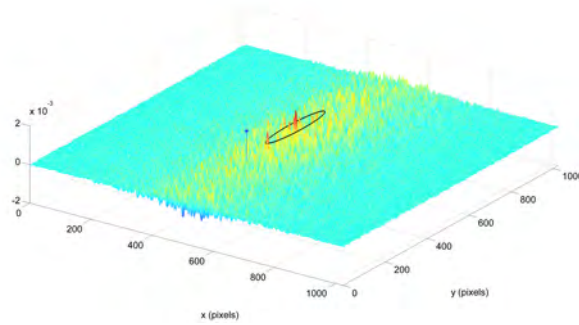
If  $K$  is too small, the obtained uncertainty measures might eventually be limited to values that do not represent the uncertainty of the main peak, leading to too optimistic uncertainty measures (as depicted in the schematic in Figure 4.4a). As can be seen, the uncertainty computed with the small  $K$  is clearly overoptimistic as only 63% of the times it encompasses the true registration result.

On the other hand, picking large  $K$  values might seem 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 lead in some cases to an over-pessimistic computation of the uncertainty (as illustrated in the schematic in Figure 4.4b). Likewise, the fact of considering a squared window may also lead to over-pessimistic estimates (Figure 4.4c). 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 two facts can explain the large values obtained for medium and large  $K$  values. While these measures explain a large percentage of the true registration results, they do so by overestimating the uncertainty, as it is indicated by their large average areas of the uncertainty ellipse. On the other hand, our proposed technique takes into account only the peaks surpassing the half power threshold and that are significant enough to condition the registration result, implicitly taking into account how they are spatially arranged. Therefore it offers the best trade-off, explaining a high percentage of the true registration results with an adjusted





**Figure 4.4:** (a) Example of a small  $K$  value that is not able to represent the uncertainty of the registration peak. (b) Example of a situation where Pfungsthorn *et al.* 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 where 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 Pfungsthorn *et al.* would use the squared window and therefore all the values in the grey area would also contribute to increase the uncertainty in the  $y$  direction.



**Figure 4.5:** Uncertainty of an unsuccessful registration not containing the true registration point. The black ellipse represents the uncertainty computed with the proposed heuristic at a confidence interval of 99%. The blue line with the star depicts the ground truth registration point.

uncertainty area. Indeed, it is important not to introduce too relaxed constraints in the graph as the optimization may quickly converge to a local minima and therefore miss the expected solution.

Note that any of the uncertainty measures reaches the theoretical 99% that should be expected. This is mainly due to the cases where the registration algorithm is unable to find a correlation point, leaving a matrix of low amplitude scattered peaks. When computing the uncertainty as described, a large value is obtained, but it is computed, and therefore centred, with respect to a peak that does not necessarily represent the registration point. Hence, in these cases, it can easily be limited out of the area where the true registration lies, as shown in Figure 4.5.

To check that effect, we recall the PSR measure described by Equation 3.5, in order to quantify



how much the peak stands with respect to the rest of the data and be able to identify those cases. The PSR has been computed for all 80000 registrations, and we have set a threshold to identify when there is not a clear dominant peak, so as to detect unsuccessful alignments. With a threshold set at PSR=20, 3208 registrations have been deemed as unsuccessful. Table 4.2 presents the same results of Table 4.1 recomputed without taking into account those registrations. It can be seen how the proposed uncertainty measure is now able to describe almost the total of the true registration results, with a further decrease on the uncertainty average area.

Method	% within $3\sigma$ bounds	Average ellipse area( $m^2$ )
<b>Pfingsthorn, K=30</b>	66%	2.42
<b>Pfingsthorn, K=60</b>	92%	26.03
<b>Pfingsthorn, K=120</b>	96%	194.86
<b>Proposed</b>	95%	1.37

**Table 4.2:** Evaluation of uncertainty measures only for successful registrations. Results of Table 4.1 computed only for registrations leading to a PSR greater than 20.

It is worth highlighting that the heuristic of Pfingsthorn *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 being easier to find a fix K parameter that can work in all cases, and resulting in 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. However, only those registrations that are deemed as successful according to a fixed threshold in the PSR measure will be input to the graph.

### 4.3.2 Navigation Constraints

Similar to registration constraints, odometry constraints computed from navigation data will have also an associated  $3 \times 3$  diagonal information matrix with  $\sigma^{-2}$  values for each measured term based on the expected measurement noise. These uncertainty terms will usually depend on the type and accuracy of the navigation instruments mounted in the vehicle (Doppler Velocity Log (DVL), fibre-optic gyro, magnetometer, etc.). For their estimation, we take as a basis the values provided by the instrument manufacturers and increment them to account for other sources of error such as deviations in the rigid-body transformations that map all sensors to the same reference point.

## 4.4 Graph Construction

Once it is clear how each type of constraint uncertainty is estimated, we need to clarify the strategy of selecting which edges will be computed and introduced in the pose graph. However, the process of building the graph will be different if we consider an offline approach, where all FLS images (and possibly navigation data) are available from the start of the processing, or if our purpose is to perform real-time mosaicing where new data is arriving at each time step. Next two sections describe the construction of the pose-graph for each approach.

### 4.4.1 Offline approach

The offline strategy contemplates the global alignment problem starting with a dataset of acquired FLS images  $FLS_{t=1..T}$ , and possibly navigation data  $NAV_{t=1..T}$ . For the sake of simplicity we assume here that  $FLS_t$  and  $NAV_t$  are paired so that both have been acquired at the time  $t$ . Note that in practice, it is probable that the acquisition frequency of FLS and navigation data do not match, and thus navigation data must be sampled/interpolated in the arrival time of the closest FLS frame. In order to compute putative edges of the described pose graph, we need to select a list of candidate frames for registration. The most naive criteria is to attempt the registration of all against all. However, this would imply a large number of unnecessary computations and consequently an excessive amount of time. For a dataset composed of  $T = 2000$  frames, and considering a registration time of 60 ms, it would take around 3 full days to compute all the registrations. Even for an offline mosaicing application this strategy is not admissible. In order 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 non-consecutive overlapping images, it is necessary to first infer the path topology.

In the absence of other sensor data, the path topology can be inferred by using the registrations of consecutive images. We compute the registrations of each frame  $FLS_t$  not only with its predecessor  $FLS_{t-1}$  but with several of their neighbor frames by establishing a fixed window  $w$  around the current sequence position:  $FLS_t, FLS_{t-1}, \dots, FLS_{t-w}$ . From that, we obtain the corresponding registration measurements  $z_{t-1,t}, \dots, z_{t-w,t}$  and information matrices  $\Omega_{t-1,t}, \dots, \Omega_{t-w,t}$ , that will form the graph edges  $e_{t-1,t}, e_{t-2,t}, \dots, e_{t-w,t}$ . The size  $w$  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 neighbour 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. In order to avoid introducing erroneous constraints in the initial graph, we discard the registrations that have been identified as unsuccessful according to an established threshold on the PSR measure, as explained in previous section. Hence, each graph vertex  $v_t$  is initialized by compounding the consecutive registrations:  $v_t = z_{1,2} \oplus z_{2,3} \oplus \dots, z_{t-1,t}$ , or, in a more generic way, by compounding along the largest connected path of the graph, as there might be the case where the registration between

two immediately consecutive images cannot be established but they can still be linked through a neighbour frame. The set of initial vertices, together with the constraints established within neighbor frames, are fed to the graph optimization back-end and an initial estimation of the path  $v_1, v_2, \dots, v_T$  is obtained. Otherwise, if navigation data is available, the initial vertex positions can be determined by compounding the odometry increments estimated from the dead-reckoning navigation:  $v_t = z_{1,2} \oplus z_{2,3} \oplus \dots, z_{t-1,t}$ , where  $z_{t-1,t} = \text{NAV}_t \ominus \text{NAV}_{t-1}$ . Note that initializing the vertices only from odometry constraints will provide a zero error solution in this initial graph optimization as the system will not be overconstrained. We can also consider merging the two types of constraints thus adding also the registration links obtained within the window of neighbour frames to estimate the initial path. Similarly, if navigation sensors provide redundancy in some DOF, extra edges could be used so that the pose graph optimizer minimizes the overall squared error in the discrepancy between measurements.

With this initial guess, putative loop closures can be identified according to the spatial arrangement of the image's positions. Hence, for each vertex  $v_t$  we will build a list of loop closure candidate frames. We will identify  $v_i$  as a candidate frame for  $v_t$  under the following criteria:

$$\begin{cases} \text{distance}(v_i, v_t) < r \\ \text{angle}(v_i, v_t) \in \left\{ -\frac{FoV}{2} : \frac{FoV}{2} \right\} \mid \text{angle}(v_i, v_t) \in \left\{ -\frac{FoV}{2} + 180 : 180 + \frac{FoV}{2} \right\} \\ v_i \notin \{v_t, v_{t+1}, \dots, v_{t+w}\} \end{cases} \quad (4.5)$$

Given that the initial path may contain some drift, we do not look for exact image overlaps but for vertices whose center is within a distance radius  $r$  of the current vertex. Furthermore, the orientation difference between  $v_t$  and the candidate frame must fall within the limits  $\left\{ -\frac{FoV}{2} : \frac{FoV}{2} \right\}$  or  $\left\{ -\frac{FoV}{2} + 180 : 180 + \frac{FoV}{2} \right\}$ . In this way, we avoid selecting as candidate pairs those frames that, even being potentially overlapping, cannot be registered given the implicit restrictions of our rotation estimation algorithm. Finally, we remove from the list the vertices whose registration has already been attempted during the initial path computation as they were part of the window of neighbour frames.

Once all the candidate pairs are identified, they are fed to the registration algorithm described in Section 3.2. The registration module returns the measurement and its associated uncertainty for each candidate pair but again only those identified as successful (according to an established threshold on the PSR measure) will be introduced as constraints in the graph. Finally, the graph back-end will run the optimization on the graph and the final set  $v_1, v_2, \dots, v_T$  of vertex locations (i.e., absolute image poses) will be obtained.

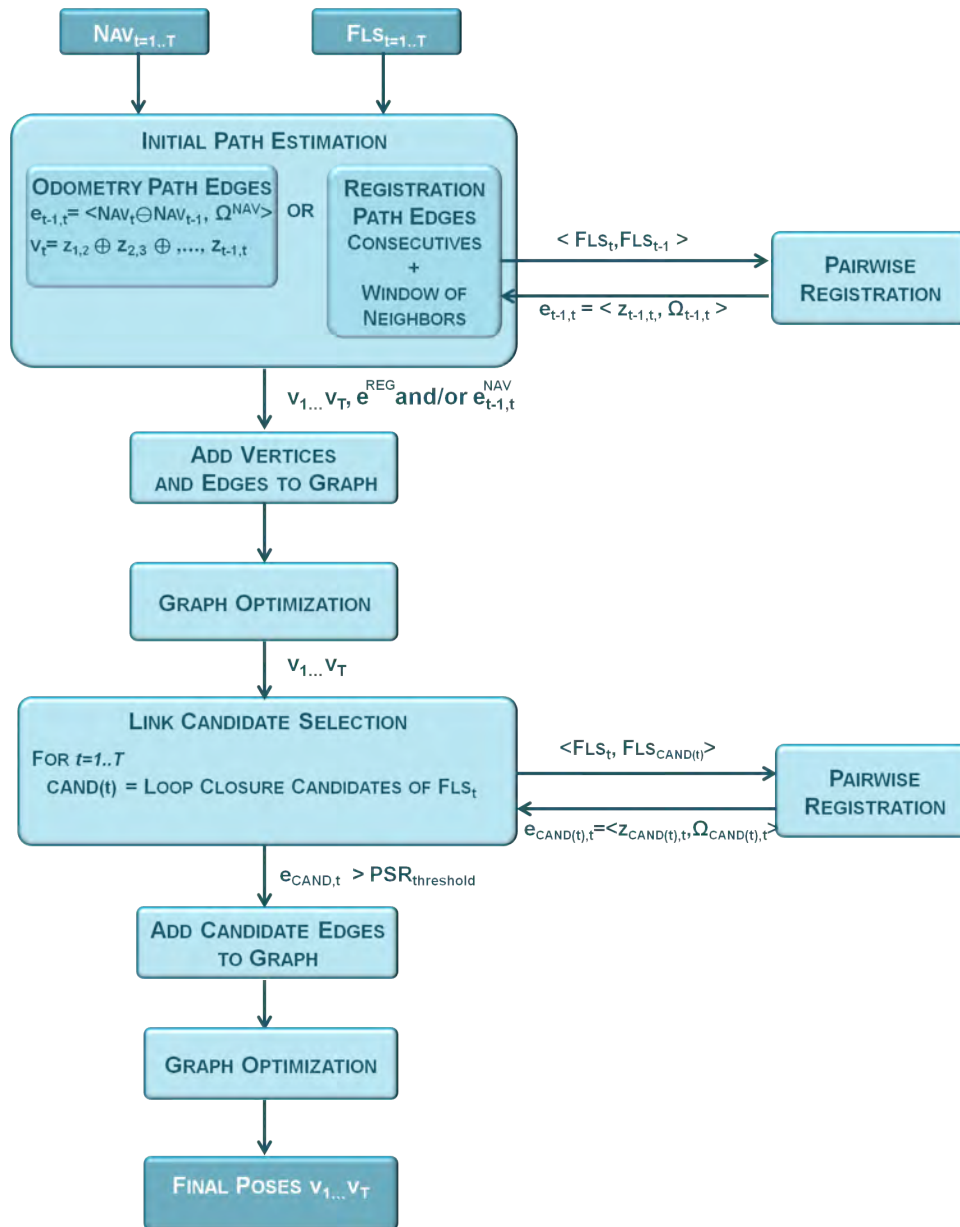


Figure 4.6: Flow diagram of the global alignment for the offline approach.

#### 4.4.2 Real-time approach

The global alignment under a real-time approach has three main differences compared to working offline:

- At a time  $t$  only the predecessor frames  $FLS_t, FLS_{t-1}, \dots, FLS_1$  are available.
- There is a limited number of registrations that can be attempted per frame in order to keep real-time requirements.
- Optimization must be performed incrementally, so as to have an updated map of absolute

poses on which to refresh the rendered mosaic.

Thus, the proposed pipeline must deal with every incoming frame, determine links with previous frames in the graph, and optimize the global position of all frames while still maintaining real-time operation.

On the arrival of a new frame  $FLS_t$ , the corresponding navigation data  $NAV_t$  are utilized to compute the odometry displacement  $z_{t-1,t}$  between the current pose and the previous one. A new vertex  $v_t$  is created and added to the graph by composing the previous vertex  $v_{t-1}$  and the computed odometry  $z_{t-1,t}$ , with the latter also added as an edge to the graph. Notice that if navigation data were not available the algorithm could initialize the vertex using the information of consecutive frame registrations. However, in the case that a given image could not be registered with its predecessors the graph would become fragmented. For that reason, in an online approach, navigation is the preferred source for the initialization of new vertices as it will always provide an odometry estimate.

Next, the received sonar frame is used as a reference frame to compute links with frames that have been already incorporated to the graph. The criteria for selecting the candidate frames is similar to that explained for the offline case, however we need to guarantee that the number of attempted registrations will be computed before the arrival of a new frame. To that end the maximum number of possible registrations  $R_{MAX}$  is calculated as:

$$R_{MAX} = \frac{1}{fps \text{ (Hz)} \cdot t_r \text{ (s)}} \quad (4.6)$$

where  $fps$  is the sonar acquisition frame rate and  $t_r$  is the time necessary to compute one pairwise registration. From the total number of possible registrations per frame  $R_{MAX}$  two different types of candidates are selected: preceding frames and temporally-far frames. The frames that are more likely to be successfully registered with a given frame are those immediately preceding the current frame. Therefore a window of frames behind the current sequence position is always included in the list of candidate pairs. However, it is also of interest to achieve registrations with temporally farther frames in order to establish loop closures. Thus, a compromise should be established in order to decide how many of each type are included in the candidate list, so that both types of registrations add to a total of  $R_{MAX}$ . Potential overlapping frames that are not temporally close can be identified according to the spatial arrangement of the already optimized vertices following the criteria of Equation 4.5. However, it is still possible that there may be too many frames satisfying the described criteria, and therefore a limited number of temporally-far frames must be selected to conform to the real-time constraint. These frames are chosen by random selection across all candidates satisfying Equation 4.5. Better selection strategies to find loop closures could be applied, for example, by taking into account the uncertainty of the vertices and selecting those edges that would lead to a greater decrease in the position uncertainty. However, the employed graph optimization back-end does not retrieve the uncertainty of the vertices (only

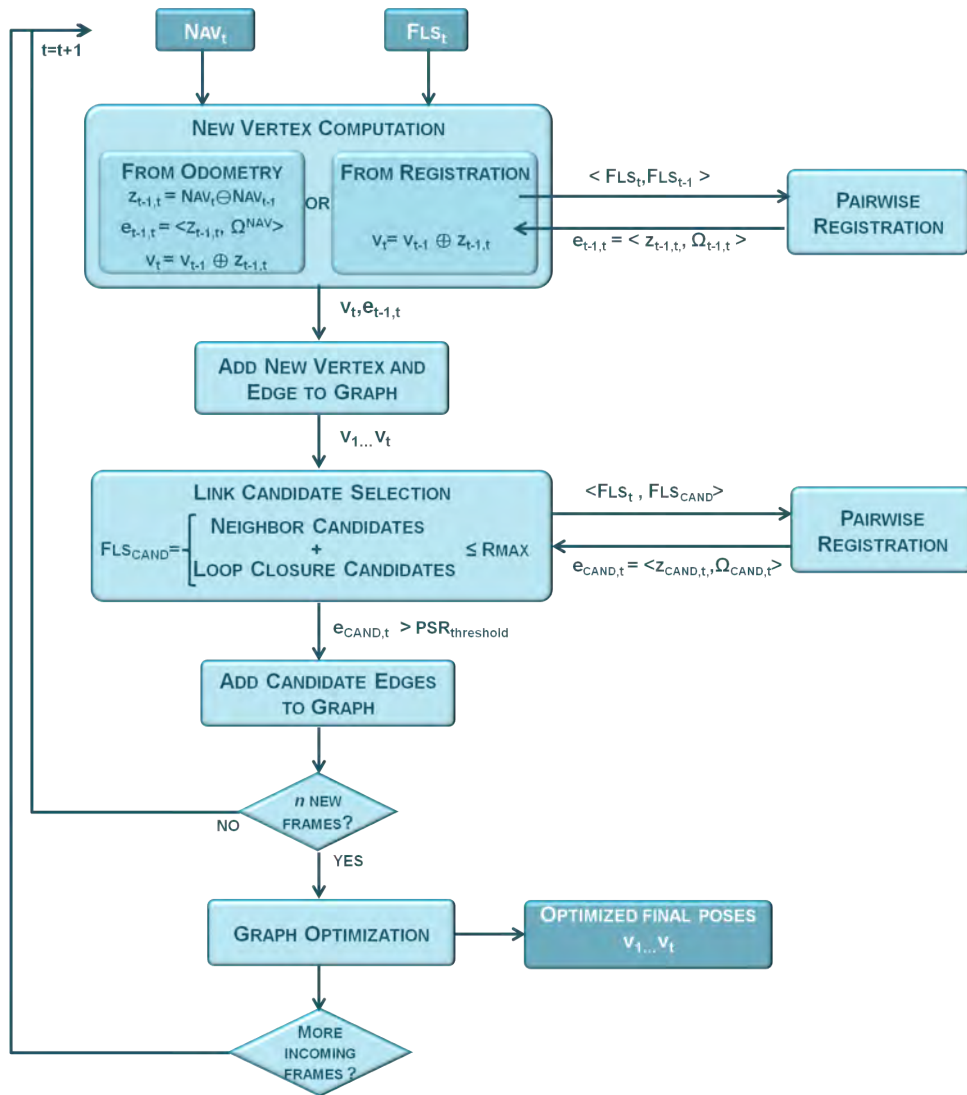


Figure 4.7: Flow diagram of the global alignment for the online approach.

that of the edges), and its computation is indeed not trivial. Therefore, we leave this possibility as a future work and we adopt the described solution.

Then, at each time step, we identify from the  $R_{MAX}$  registrations the ones that are deemed as successful and we incorporate them as edges into the graph. This process is repeated until  $n$  vertices are added to the graph, at which moment the optimization is performed yielding the optimized poses until the current time step  $t$ . Therefore, at each  $n$  time steps the map of absolute poses will be updated, being available to the renderer module to refresh the mosaic visualization.

Two important aspects must be highlighted here in regards to the real-time capability. The first is the importance of using a registration algorithm that can cope with the majority of the image artifacts without requiring the preprocessing of the frames. This turns into more time to attempt a higher number of registrations and therefore increase the possibility of including more

graph constraints.

Second, the online workflow presented in this section, does not guarantee that a real-time operation will always be possible as there exist some scalability bounds. Despite the efficiency of smoothing approaches, they are not constant time when closing large loops [Kaess et al.2012], so eventually the real time capability of the optimization would drop off.

Related also to scalability, there is the issue of frame storage and retrieval for long missions. In order to hypothesize loop closures, previous frames must be available along the vehicle trajectory. For graphs of the order of thousands of frames, where all frames can be kept in RAM, the system can hold real time capabilities as will be demonstrated by the experiment of Section 6.4. However, despite the low resolution of FLS images, holding all data will become at some point unfeasible. Other solutions such as conveniently splitting the map in submosaics and simplifying the graph over time should be explored in this direction.

# 5

## MOSAIC RENDERING

---

The last step of the mosaicing pipeline is the rendering of the image mosaic. After the global alignment step covered on the previous chapter the poses of the sonar images in a global reference frame are obtained. We can then build the corresponding absolute homographies and map the images on the mosaic plane. However, as the content of multiple images will overlap at a given position, a strategy is required to deal with the combination of the pixel intensities. In this chapter we study the problem of fusing the different FLS frames while minimizing the photometrical differences between them in order to generate a mosaic of smooth and continuous appearance. We will first lay out the differences of the problem at hand with respect to the problem of blending optical images. Then, we will propose a compendium of strategies that are targeted to address the photometrical irregularities that can be present when mosaicing FLS images. Finally we also provide insights on the SNR and resolution improvement of the generated mosaics.



## 5.1 Blending of acoustic images

Usually, when video frames are projected on a photomosaic, the seams along the image boundaries become noticeable due to photometrical differences between the individual frames. The same occurs with FLS images, often in a more accentuated manner due to multiple and strong photometric artifacts. Thus, it is essential to perform one last step to give a continuous and uniform appearance in the form of a single large mosaic. This is achieved by means of image blending techniques. It is worth highlighting that generating the sonar mosaic with a smooth and continuous appearance is not merely driven by aesthetic purposes but it is essential for a more effective interpretation. If images are plotted in a last-on-top fashion, without any fusion mechanism, the photometric artifacts affecting the FLS frames can cause severe degradation of the final mosaic up to the point where understanding the imaged scene becomes impossible.

The basic principles of image blending were established four decades ago [Burt and Adelson1983] and the topic has been extensively studied in the field of optical imaging, including underwater environments [Prados et al.2012]. There is a wide variety of image blending techniques in the literature, but at a high level two main approaches can be distinguished. On one hand, we have transition smoothing methods (also known as or alpha blending methods [Porter and Duff1984] or feathering [Uyttendaele et al.2001]) which attempt to minimize the visibility of the image boundaries by smoothing the common overlapping region of the stitched images. On the other hand, there are optimal seam finding methods [Fan Gu and Rzhano2006, Gracias et al.2009] which attempt to find the optimal location to place a cut along the two images so that it minimizes the photometrical and geometrical changes between them. Furthermore, there are also hybrid techniques which take advantage of the benefits of each approach [Prados et al.2012].

Hence, regardless of the particular technique, optical blending generally deals with a low number of images at a given position (most of the times pairwise) and treats only their intersecting region. However, blending a FLS mosaic requires dealing with multiple overlapping images involving high overlap percentages, which prevents us to leverage from traditional blending techniques designed for video images. Specially 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 by an averaging process, as will be described in Section 5.2. This is again opposed to other approaches typically adopted on optical mosaicing such as trying to select only the best image for a given location. Thus, to blend FLS mosaics it is necessary to deal not only with the seam areas, but with the fusion of the whole image content.

The state of the art in image blending does not include precise solutions to deal with acoustic images and, in fact, little work can be found in the literature regarding sonar image blending. In [Coiras et al.2004] side-scan sonar data is blended using a wavelet-based technique that

allows to select which kind of features are emphasized in the final mosaic. Kim et al. [Kim et al.2008] proposed a probabilistic approach in the context of a superresolution technique for FLS frames. They model the blending problem of fusing a low-resolution image into a high-resolution one in terms of a conditional distribution with constraints imposed by the illumination profile of the frames so as to maximize the SNR of the resulting image. In our initial work on FLS mosaicing [Hurtós et al.2012], results have been rendered by averaging the intensities of all overlapping sonar frames at every mosaic pixel. In other research reporting FLS mosaic results, such as in [Negahdaripour et al.2005] or [Negahdaripour et al.2011], the image blending method is not specified and we presume pixel averaging is used as well. However, as these mosaics are composed only of dozens of frames with very homogeneous intensity profiles, even a last-on-top strategy could have been used in this context. In general, none of the existing strategies deals with the particular issues of FLS images and the photometric artifacts that can arise from its imaging conditions. Moreover, as previous works have only reported small single-transect mosaics, there are some particular challenges in blending large multiple-track mosaics that need to be specifically addressed to obtain a proper sonar blended mosaic.

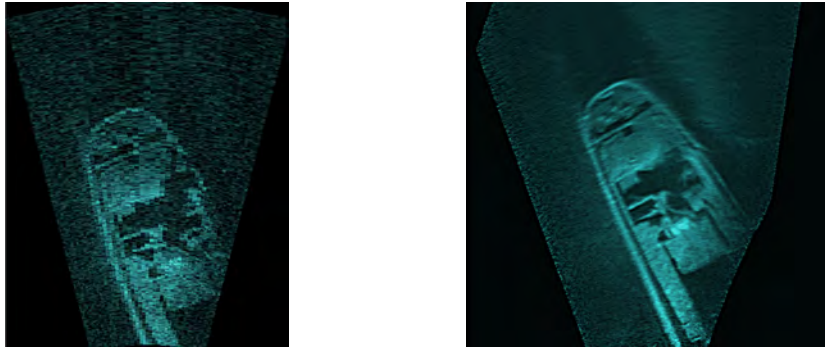
Next section will analyze the basic blending through averaging, which is the basis of our proposed image fusion mechanism. Then, a compendium of strategies are provided to minimize the different photometric artifacts and achieve a visually pleasant result.

## 5.2 Intensity averaging blending

A simple but effective strategy to blend a FLS mosaic is to perform an average of the intensities that are mapped to the same pixel location. Assuming that a correct image alignment has been found, averaging the overlapping image intensities yields the denoising of the final mosaic.

Thus, by adopting an averaging approach, the resulting mosaic will have a better SNR compared to a single image frame (see Figure 5.1). Ideally, by averaging, the reduction of the noise (and therefore the improvement of the SNR) is proportional to the squared root of the number of averaged samples. Then, under the assumption of additive uncorrelated noise, a mosaic would have an overall SNR improvement proportional to the mean of the square roots of the number of images at each pixel location. However, we must highlight 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, which serves also as a visual indicator of the mosaic's consistency. Further advantages are that, given its simplicity, the blending is suitable to be implemented in real time. Assuming that the full mosaic can be stored in memory, the mean pixel values can be efficiently recomputed in an incremental way for the area where each incoming image is mapped. The ability



**Figure 5.1:** 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.

to not only compute but render a mosaic in real-time can be of interest in many situations. A clear example would be the generation of an online map for improving the situational awareness of a ROV pilot as he drives the vehicle. However, examples can also be found in the context of an AUV application such as for instance a search and reacquisition task without the need of resurfacing. We can envisage a mission where an AUV surveys an area, sends the map to the surface and from there an expert identifies the target to explore, and sends its position back to the AUV to command the inspection. Since AUV-surface communications are subject to the limited bandwidth of acoustic modems, transmitting the mosaic instead of the individual frames offers a compact and feasible way of performing such task. Moreover, the benefit of the denoising effect becomes even more relevant in that case since an image with increased SNR will typically be better compressed and thus will be more efficiently transmitted.

Although the averaging approach is a good starting point and may give satisfactory results in some imaging configurations, it suffers from several problems. Averaging tends to diminish details in those places where there is a large number of frame contributions. A fade-out of the mosaic content can occur as a consequence of non-uniform intensity distribution in the images or when averaging frames that have blind (i.e., black) areas, usually due to non-proper imaging configurations. Besides, image boundaries or seams between different transects can become noticeable due to a non-constant number of overlapping images, especially in datasets with multiple tracklines, rotations or non-constant vehicle speed.

### 5.3 Improving the blending of FLS mosaics

The proposed methodology consists of a set of strategies to address the problems explained in the previous section associated with FLS image composition. Our approach takes as a basis the fusion by intensity averaging and incorporates strategies to correct for the different artifacts and modify the number and/or the intensity of the averaged pixels for the final image composition.

It is important to stress that the proposed strategies concentrate on solving the different photometric artifacts that can arise both at individual frame level and at global mosaic scale but does not focus in possible problems caused by geometric registration errors. Therefore, a correct registration is assumed from previous steps, otherwise the averaging principle would smear the content of the misregistered area. In the same way, other geometric issues such as object shadow alterations due to the viewpoint change are handled implicitly by the averaging principle. For instance, when imaging a protruding object while navigating over it, its shadow gets shortened as the sonar becomes closer to the object. The final shadow representation in the mosaic will then be the mean of all shadow positions yielding an intermediate solution which we consider to be a reasonable description of the scene.

Each of the strategies presented here can be enabled or disabled in the blending stage according to the characteristics of the dataset. Therefore a dataset gathered in ideal conditions (i.e., with a sonar that would not present inhomogeneous insonification patterns, with the proper altitude, tilt and range settings, imaging a planar scene and performing just a single trackline at constant speed in order to keep a uniform number of overlapped images) would be blended through a standard intensity averaging only benefiting from the local contrast enhancement step to emphasize its features.

It is worth noting that the proposed blending is designed to work in an offline fashion as it requires using all gathered frames with the aim of producing a final high-quality map of the inspected area. On the other hand, for real-time mosaicing the simple averaging blending is the adopted solution.

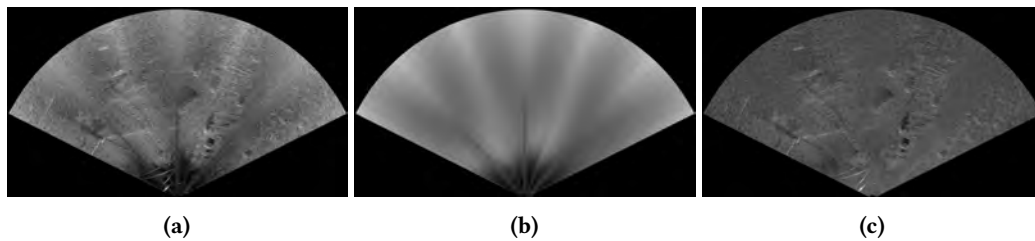
### 5.3.1 Individual image pre-processing

The described photometric artifacts that occur at image level (see Section 2.3) can affect the global appearance of the mosaic composition. Hence, it is important to pre-process individual sonar frames to correct for some possible irregularities. Note that, unless otherwise stated, this pre-processing is only necessary for the blending stage. The registration method covered in Chapter 3 is generally able to perform registration under the effect of these artifacts and therefore reduces the need for preprocessing the images before the registration.

#### 5.3.1.1 Inhomogeneous insonification pattern correction

Some sonar models show evidence of non-uniform insonification patterns due to the different sensitivity of the transducers across the field of view (Fig. 5.2a). If a sufficient number of images is available, the underlying illumination profile can be computed by averaging all the dataset frames (Fig. 5.2b). Then, the illumination profile can be compensated in the original image thus yielding a pattern-free image (Fig. 5.2c). If the pattern is strong, this step should be performed earlier in the mosaicing pipeline (i.e., prior to the pairwise registration of the sonar images) since

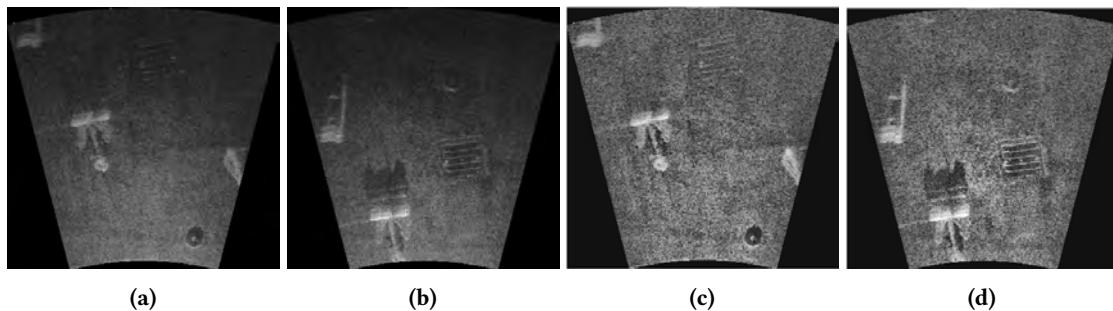
its presence may influence the registration result.



**Figure 5.2:** Correction of non-inhomogeneous insonification pattern. (a) Original frame. (b) Estimated illumination pattern. (c) Corrected frame.

### 5.3.1.2 Contrast Limited Adaptive Histogram Equalization (CLAHE)

Besides non-uniform insonification related to the sensor's hardware, FLS data can exhibit other non-constant illumination patterns. Due to non-compensated transmission loss, imaging configuration and/or terrain curvature, the images can exhibit weaker backscattered intensities in some areas (e.g., weaker intensities further away from the sonar origin). This results in considerable intensity offsets when registering images that insonify the same portion of the scene but from different locations (Figure 5.3a, 5.3b), turning into a visible attenuation when blending the mosaic.



**Figure 5.3:** Image equalisation and contrast enhancement. (a)-(b) Example of two frames imaging the same area from two different viewpoints. Notice the difference in intensities around the grid in the center. (c)-(d) Same frames preprocessed with CLAHE. The images present a uniform distribution that allows to merge them without attenuating the final mosaic. Notice also that the local contrast is preserved, emphasizing the scene features.

To deal with this, we first equalize the intensity histograms of the sonar frames so as to match a uniform distribution and minimize the intensity offsets on the registered areas. To this end, we employ the CLAHE technique [Pizer et al.1987] whose advantages are twofold: first, it equalizes the images limiting the noise in the areas that are more homogeneous by setting a clip limit on the histogram equalization; second, it locally enhances the contrast of the images alleviating the low SNR that characterizes FLS images. Although, as stated before, the SNR is greatly enhanced by the averaging nature of the blending, a local contrast enhancement can help to further emphasize the scene features.

Note that this procedure does not preserve the true reflectivity values of the scene. However, we believe that for inspection purposes, it is more important to obtain a continuous and smooth representation that emphasizes the features and facilitates a better scene interpretation rather than preserving the true scene reflectivity.

### 5.3.1.3 Masking out blind regions

Extreme cases of non-uniform intensities across the images are those situations where either an inappropriate imaging configuration or significant relief variations introduce blind regions in the sonar frames. Even when working with equalized images, those blind regions have a negative impact on the final blending. Since they do not contain information at all, they cause the actual scene content to fade out when they are averaged with other images.

Our strategy for those cases is to compute a saliency mask for each frame (Fig. 5.4), which will be used to mask out the blind regions when performing the fusion by averaging. The mask  $m$  is obtained by applying standard deviation to local neighbourhoods:

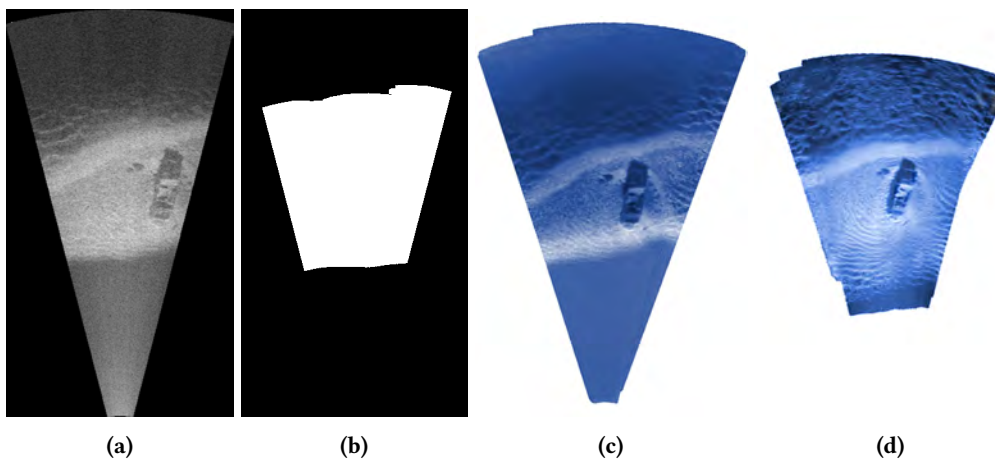
$$m(u, v) = \sqrt{\frac{\sum (i(x, y) - \bar{i}(x, y))^2}{n - 1}}, \quad (5.1)$$

where  $n$  is the number of pixels in the neighbourhood,  $i(x, y)$  is the intensity of the pixel under consideration and  $\bar{i}(x, y)$  is the mean of all neighbourhood pixels. The shape and size of the local neighbourhood are parameters that can be adjusted so as to take into account the standard deviation generated by the residual noise of the images. This standard deviation filter acts as a texture classifier. The blind regions of the image, which are characterized by the lack of backscattered intensities report low values. On the other hand, scene backscattered intensities generate higher filter responses. Hence, a threshold is set to segment both type of regions. To avoid the selection of image parts corresponding to homogeneous areas on the imaged surface (such as in a flat sandy bottom) an additional threshold on the mean value of the pixels is imposed. Therefore, blind areas will be homogeneous areas with low response of the standard deviation filter that are characterized by a low average intensity. Finally, to avoid including areas of internal shadows that can also conform to the previous description, we check for connectivity with the image footprint boundaries.

## 5.3.2 Global mosaic blending

Despite the illumination corrections performed at individual frame level, the fusion of images from different tracklines will unavoidably create noticeable seams along the tracks due to the presence of a higher number of image contributions in the overlapping area (Figure 5.5a). If the tracks are combined along-range, seams may be also noticeable as a consequence of merging two





**Figure 5.4:** Example of dataset frames affected by blind areas and varying illumination depending on the imaging viewpoint. (a) Example of an original image with blind areas. (b) Corresponding mask (black pixels will not be taken into account in the blending). (c) Mosaic after standard intensity averaging. (d) Mosaic blended by applying CLAHE and the masking of blind regions. Note that, opposed to (c), the sand ripples above and below the center target can be clearly appreciated.

different image quality areas (low/high resolution).

To reduce these artifacts, we compute an overlap map that reflects the number of images projected at each pixel location, taking into account the possible masks that might have been computed previously if the images contain blind regions. In the presence of multiple tracklines, the intersection area will present a significantly higher number of overlapped images compared to their surroundings (Fig. 5.5b).

To avoid these artifacts we propose a mechanism consisting of three main steps: clipping, smoothing and selection. First, the number of overlapping images are clipped to a threshold thus reducing the range of possible different overlaps. While it helps to reduce the overload of pixel contributions at a given location it is also of interest to keep a significant number of overlapped images to diminish the noise of the final mosaic. A trade off solution consists of cutting up to the mean of the overlap map values. Second, the new overlap map is smoothed with a Gaussian kernel to avoid sharp transitions caused by a different number of pixel contributions. A normalization is required so as to avoid any new computed overlap to exceed the number of actual overlapping images. Finally, the mosaic is blended by averaging the number of pixels indicated by the new overlap map. To select from the images that are projected to a given pixel which ones will be discarded and which ones will be taken into account on the final averaging, the following procedure is implemented: for each pixel of the final mosaic we store the list of all values that are projected to that location together with a weight that reflects its position in its original frame. A weighting mask is used to reflect the location, and therefore the resolution, of each pixel (from higher to lower as measurement sparseness increases with the range on Cartesian space). In this way, candidate pixels are sorted according to their weight and the first  $n$  ones of higher weight (being  $n$  the number of overlapping pixels in the newly computed overlap map)

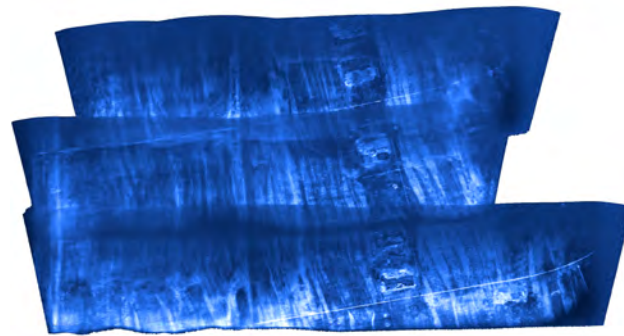
are used to compute the final pixel intensity by averaging. In this way we give priority to those frames that depict the region with higher resolution.

## 5.4 Resolution enhancement

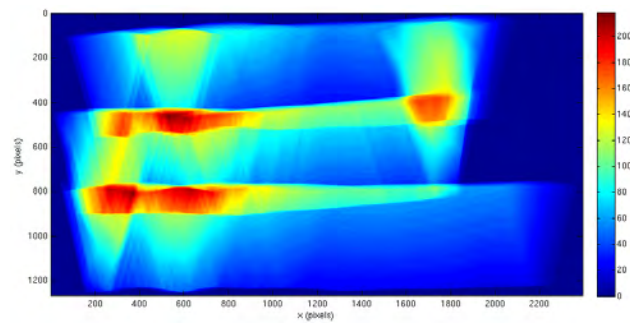
Apart from the improvements in SNR that we achieve as a result of the averaging blending, the spatial resolution of the final mosaic can also be enhanced with respect to the original images. The actual resolution in which a mosaic is rendered is, by default, the same as that of a single frame. In this sense, a frame and a corresponding mosaic covering the same area will have the same amount of pixels. However, a single FLS image has a much lower *apparent* resolution, as a consequence of the inhomogenous mapping from polar to Cartesian coordinates that assigns the same pixel intensity to several pixels in the Cartesian domain. In the mosaic, this effect is attenuated as a consequence of the averaging of multiple frames, thus leading to an improvement of *apparent* resolution.

Nevertheless, the real resolution can be indeed increased thanks to the mosaicing process. We can take advantage of the multiple alignment of low-resolution images together with the subpixel accuracy positions obtained from the global alignment step to perform super-resolution. 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.

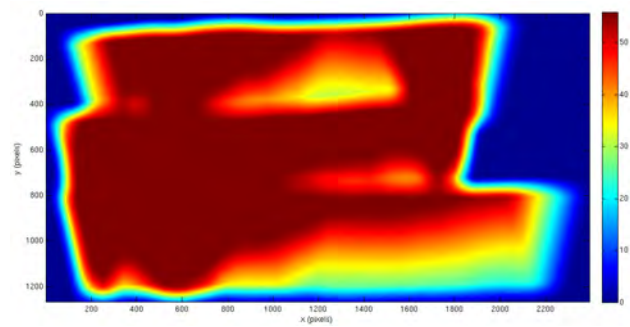




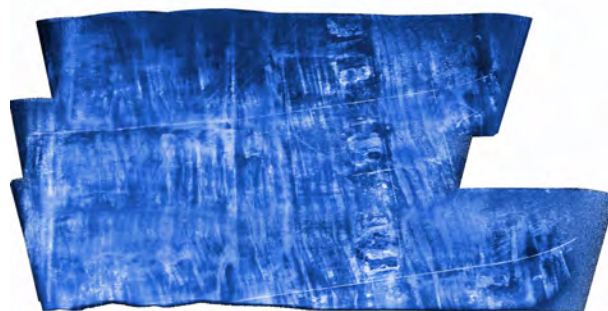
(a)



(b)



(c)



(d)

**Figure 5.5:** Mosaic blending corrections at global level. (a) Detail of a DIDSON sonar mosaic presenting three different straight tracklines. Note the visible seams at the regions of track intersection. (b) Overlap map showing large differences in the number of overlapping images across the mosaic. (c) Computed overlap map clipping the highest overlap values and applying smoothing to avoid abrupt changes. (d) Blended mosaic.

# 6

## EXPERIMENTS AND RESULTS

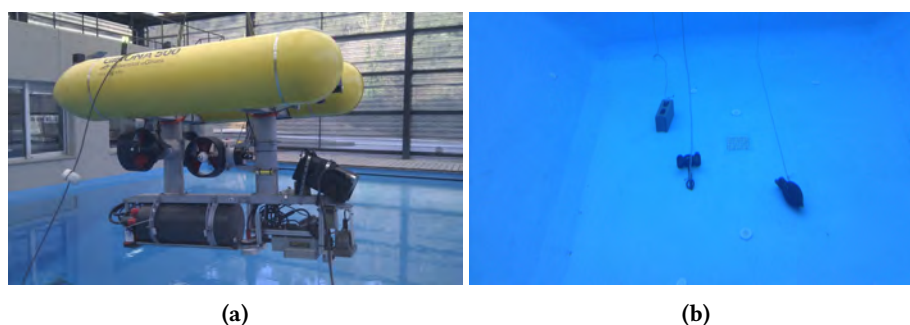
---

Once the different steps of the mosaicing system have been described and analyzed individually, this chapter presents experimental results validating the full pipeline. We first report on a small test performed inside a water tank, followed by results on real-world field applications that take place inside environments where visibility is often compromised. In order to demonstrate the applicability of the mosaicing system to varied conditions, the datasets used in this chapter include images acquired from different platforms, different sonar models and affected by diverse artifacts. We utilize each experiment to highlight different aspects of our pipeline's performance, involving mosaic consistency, accuracy or real-time capabilities. For every experiment, we report the details of the input sonar frames, the computation times and the resolution and size of the generated mosaics. All computations are performed on an Intel i7 3.4 Mhz QuadCore CPU and, unless otherwise stated, the registration algorithm is running on a single core.

## 6.1 Tank test

### 6.1.1 Experiment description

This experiment was carried out in the CIRS water tank at the University of Girona using the Girona-500 AUV [Ribas et al.2012]. The vehicle is equipped with an RDI Doppler Velocity Log and a Fiber Optic Gyro as main navigation sensors. An ARIS Explorer 3000 sonar from Sound Metrics [Sound Metrics ARIS2013] was additionally installed for this experiment (Figure 6.1a). Several objects of small size (i.e., around 50 cm) were deployed at the bottom of the tank (see Figure 6.1b). The vehicle was teleoperated over them describing a short lawn mower pattern while trying to maintain 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 deg 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 unit.



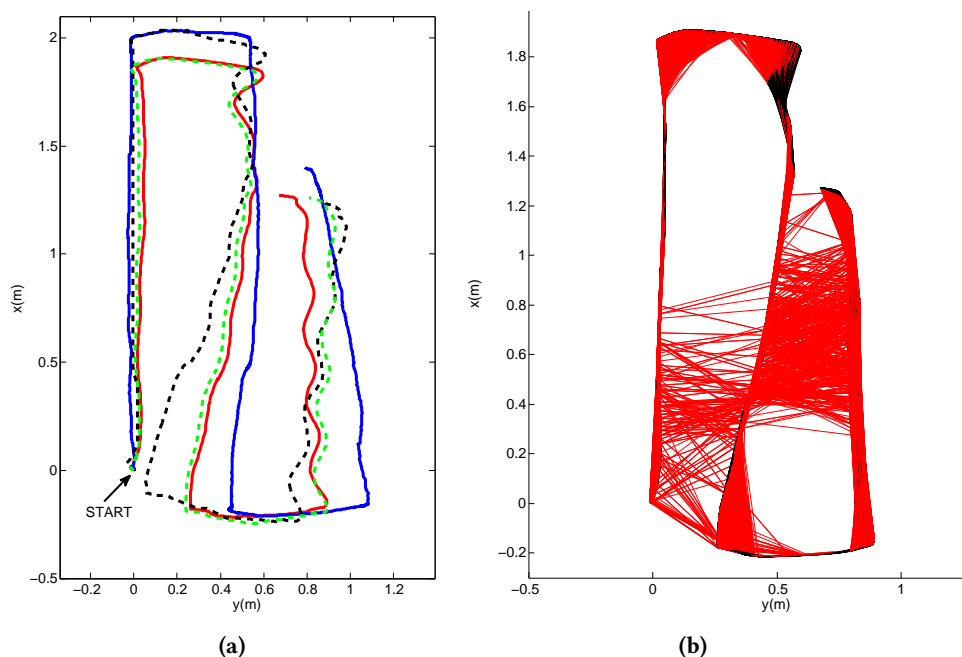
**Figure 6.1:** Tank experiment setup. (a) Girona 500 AUV equipped with ARIS FLS. (b) Objects deployed in the water tank.

TANK DATASET			Example frame
<b>Input frame</b>	<i>Size (pixels)</i>	<i>Resolution(m/pixel)</i>	
	350x274	0.0045	
<b>Computation times</b>	<i>Registration (s)</i>	<i>Optimization (s)</i>	
	2312	27.6	
<b>Final mosaic</b>	<i>Size (m)</i>	<i>Resolution (m/pixel)</i>	
	2.3x3	0.0011	

**Table 6.1:** Summary table for the Tank Test Dataset.

### 6.1.2 Results

We have proceeded with the application of the proposed mosaicing pipeline in an offline fashion over the described dataset. Figure 6.2a shows the vehicle's dead-reckoning trajectory together

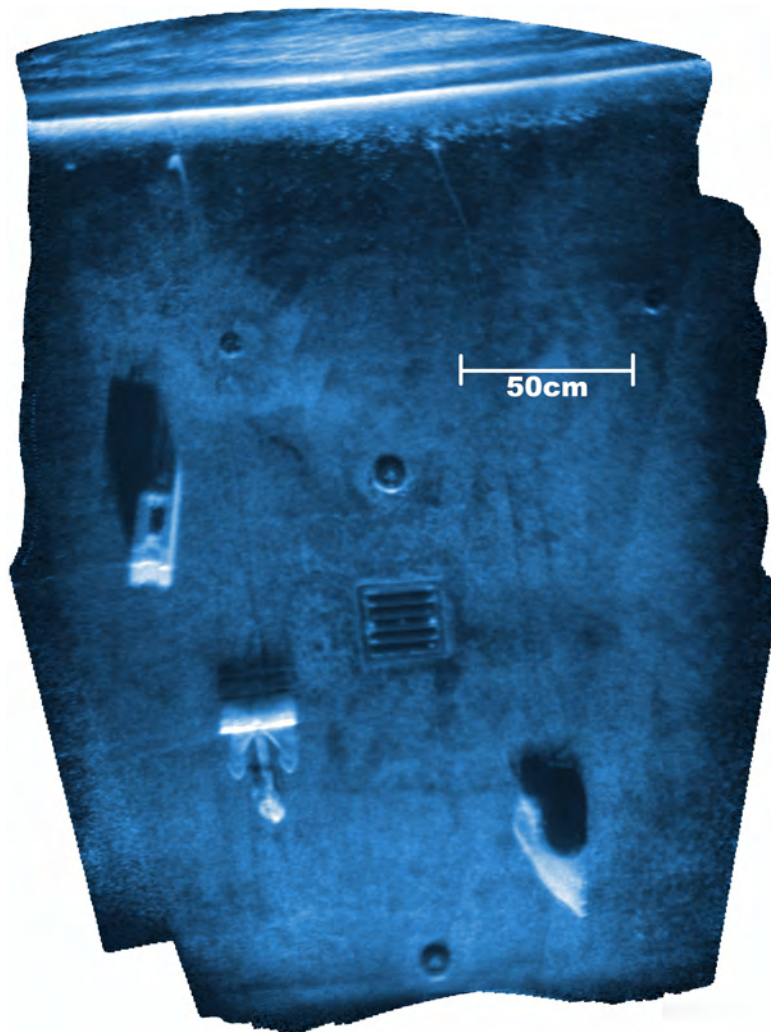


**Figure 6.2:** (a) Trajectories of the ARIS Tank experiment. Blue: Vehicle's dead-reckoning trajectory. Black-dashed: Trajectory estimation from consecutive image registrations. Green-dashed: Trajectory estimation from the consecutive constraints including a window of local neighbours. Red: Final estimated trajectory after the global alignment. (b) Final graph constraints of the ARIS Tank experiment. Black: window constraints. Red: loop-closure constraints.

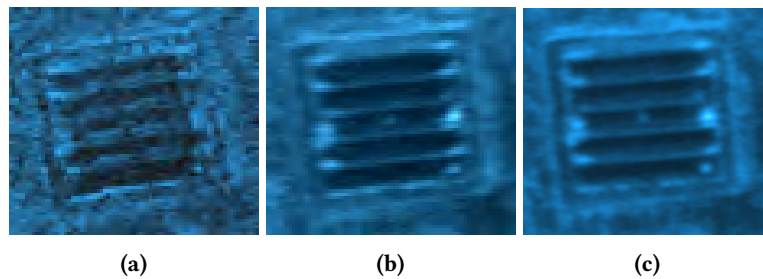
with several of the estimated trajectories. The black dashed line shows the estimated trajectory computed by concatenation of the registration constraints of consecutive image pairs. The green dashed line shows the estimated path including constraints computed within a window of 20 neighbouring frames. 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). The green path has been used as the initial guess to find loop closure hypothesis. Due to the high overlap of the sequence, the method returned 36092 potentially overlapping pairs under the requirement of 1 m distance between image centres. From these, 15473 were considered successful registrations according to strict PSR ratio of 25 and were added to the graph together with their uncertainties. Figure 6.2b depicts the final graph, where the high number of included constraints can be appreciated.

Figure 6.3 shows the obtained mosaic composed of 527 frames and rendered over an oversampled grid at 4 times the original resolution. The mosaic shows high self-consistence, testifying the accuracy of the method, and enables the identification of the small objects present in the scene: a concrete block, an anchor and an amphora as well as the grid and other details of the tank. Figure 6.4 shows a comparison of a detailed area where the SNR improvement between a single sonar frame (Figure 6.4a) and the mosaic (Figure 6.4b) can be easily appreciated. Note also the enhancement of the image when comparing the oversampled (Figure 6.4c) and non-oversampled

(Figure 6.4b) versions of the mosaic.



**Figure 6.3:** Mosaic composition with the 527 frames from the tank experiment.



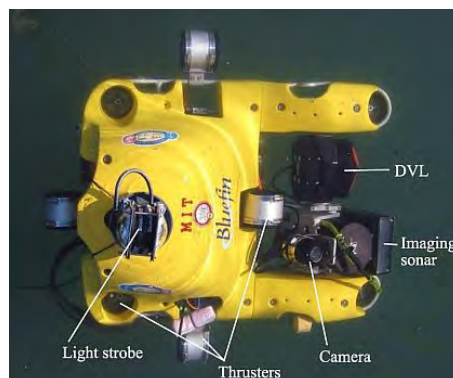
**Figure 6.4:** Detail comparison of a small area in the tank mosaic. (a) Single frame. (b) Original resolution mosaic. (c) Mosaic oversampled 4 times the original resolution. Note the improvement of the mosaic SNR with respect to the individual frame and the detail enhancement of the oversampled mosaic version.

## 6.2 Ship Hull Inspection

### 6.2.1 Experiment description

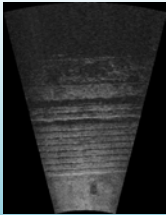
The second presented experiment is in the context of ship hull inspection. Ship hulls are routinely inspected for security reasons using divers, being a hazardous and time consuming task. Recent efforts have demonstrated the capability to navigate autonomously with respect to a ship hull using an AUV [Hover et al.2012] in order to collect visual or acoustic data of the hull in a safe and reliable manner. Given that these inspections are carried out inside harbor, where water visibility is often limited, they are a good example of target application for the proposed mapping methodology.

The dataset used in this section is courtesy of Bluefin Robotics [Bluefin Robotics Corp.2013] and was acquired with the Hovering Autonomous Underwater Vehicle (HAUV) [Vaganay et al.2005] equipped with a DIDSON sonar [Sound Metrics DIDSON2013] (Figure 6.5). The experiment was conducted on the King Triton vessel in Boston Harbor. The vehicle navigated across the bottom of the hull, maintaining a constant distance to it and covering an area of about 15 m x 6 m. The sonar was mounted on a tilt unit and was actuated throughout the experiment to adapt the images to the hull's surface. The final trajectory consists of 5 tracklines across the bottom of the hull, comprising a total of 4420 sonar images collected during 12 minutes. 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 sway DOFs, which facilitates the registration between revisited locations as the vantage point is preserved throughout the experiment.



**Figure 6.5:** HAUV vehicle with DIDSON FLS used in the ship hull experiment. Image credit: Bluefin Robotics.

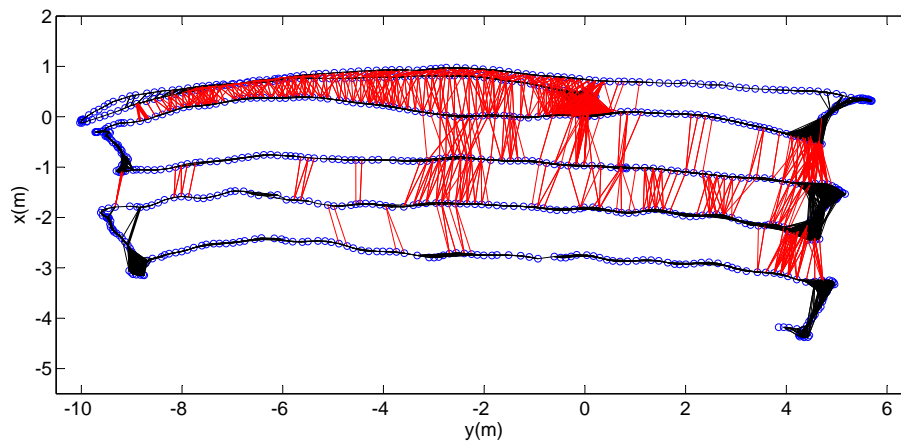


SHIP HULL DATASET			Example frame 
<b>Input frame</b>	<i>Size (pixels)</i>	<i>Resolution(m/pixel)</i>	
	350x453	0.01	
<b>Computation times</b>	<i>Registration (s)</i>	<i>Optimization (s)</i>	
	2654	31.3	
<b>Final mosaic</b>	<i>Size (m)</i>	<i>Resolution (m/pixel)</i>	
	19x9.5	0.003	

**Table 6.2:** Summary table for the ship hull dataset.

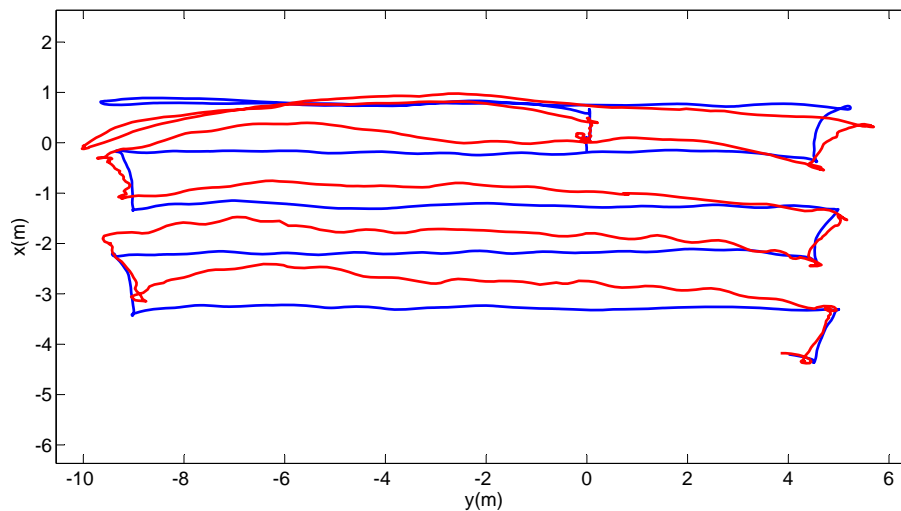
### 6.2.2 Results

Due to the high frame rate of acquisition (6 frames per second), only one out of 3 images has been considered, reducing the dataset to 1473 frames. Although available, no navigation information has been used in the global alignment stage. Following the link candidate strategy, a total of 17079 pairwise registrations have been attempted, including frames from the vicinity of the sequence and frames found in loop closure situations. From these, 8148 have been deemed correct registrations according to a PSR threshold of 20. The high number of established constraints (consecutive and non-consecutive) 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 minutes, which suggests that by applying the online pipeline with a more restrictive pruning on the attempted frames the mosaic could be constructed in real-time.



**Figure 6.6:** 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 neighbouring frames inside a window. Links in red represent constraints found in loop-closure situations.

Figure 6.6 shows the final computed trajectory from the sonar images, depicting all the links established between frames. Figure 6.7 shows the navigation trajectory (in blue) and the trajectory computed with our methodology (in red) both referenced at the sonar's origin. Unfortunately,



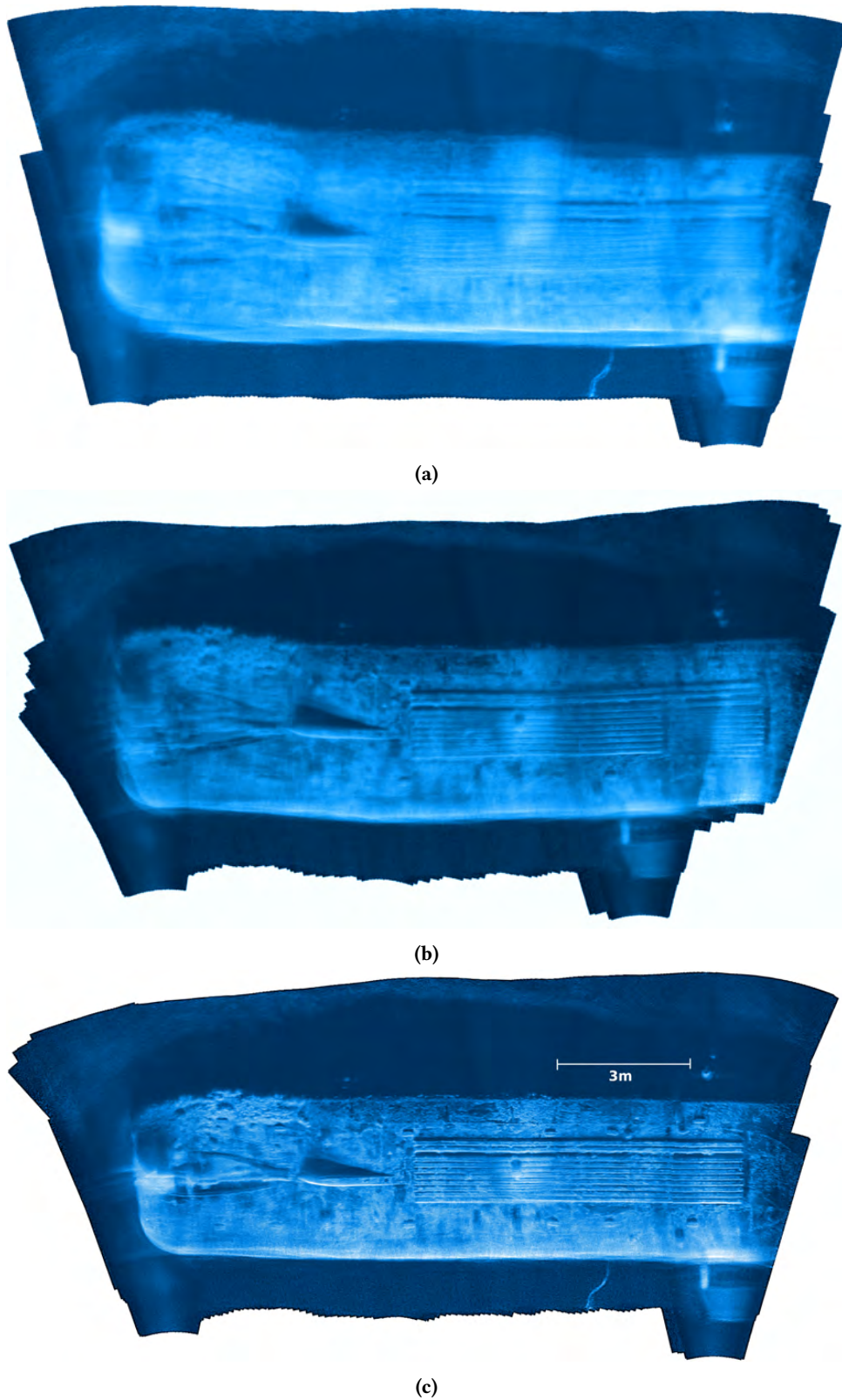
**Figure 6.7:** Trajectories of the ship hull dataset. Navigation trajectory (in blue) and estimated trajectory after the global alignment (in red).

as ground truth is not available in this dataset, we cannot provide a quantitative measure of which trajectory is closer to the reality. 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 consistent image composition. Figure 6.8a shows the mosaic over the navigation trajectory, while Figure 6.8b displays the mosaic of the estimated trajectory prior to the optimization step and Figure 6.8c shows the final obtained mosaic oversampled by a factor of 3. It can be seen that the composite image in Figure 6.8c presents a consistent overall appearance and allows the identification of the various features on the hull. Some illumination artifacts are present (especially in the lower part of the image) due to the tilt imaging angle.

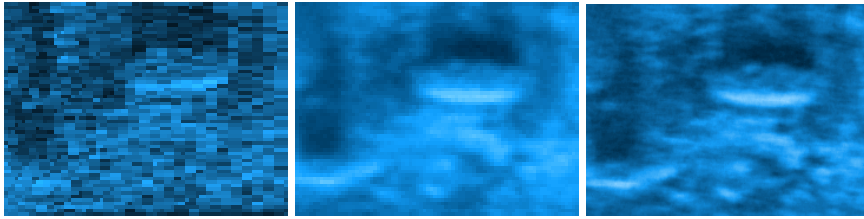
Figure 6.20 shows a comparison of a detailed area between a single sonar frame, the non-oversampled version of the mosaic, and the mosaic oversampled at 3 times the original resolution.

We would like to take profit from the fact that we have a second dataset of the same ship hull to emphasize the importance of a proper data acquisition. The second dataset is composed of 518 frames gathered along a single trackline where the entire ship hull's width is imaged within the sonar's FOV. Figure 6.10 shows a comparison of the mosaics obtained with the two datasets. As can be seen, although the overall shape and dimensions of the ship-hull agree in both mosaics, the single-pass one presents a better definition of the hull details (Figure 6.10b). This is due to the differences on the input data, that was acquired under different conditions. The first dataset presents severe sonar tilt variations along the trajectory, ranging from values as different as from 14 to 33 degrees. Although this does not constitute a problem to register the images it does slightly alter the shadows and appearance of the features along the trajectory leading to a less defined composition. Besides, the dataset undergoes also roll motions spanning values over 15 degrees thus contributing to add small misalignments. And finally, several spurious artifacts arise during the acquisition: in the first trackline there are strong reflections from the water surface and in

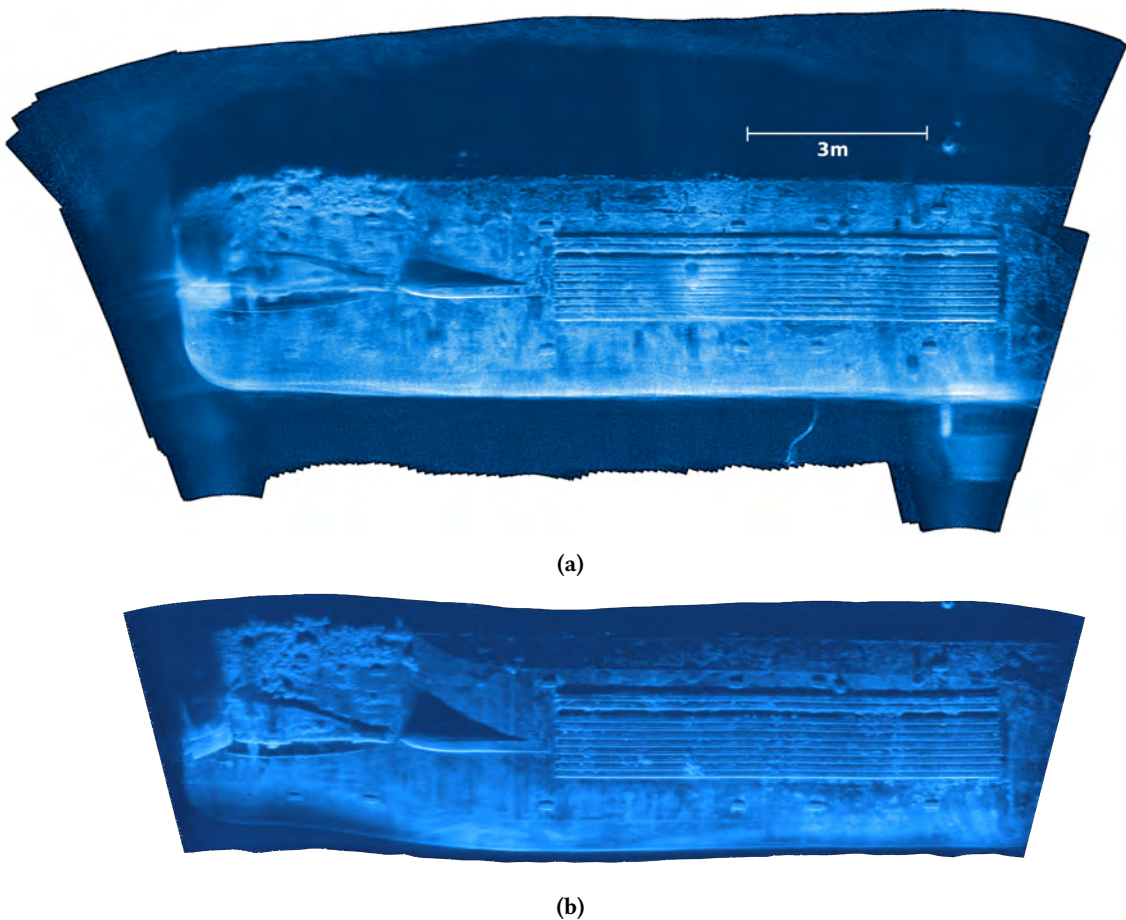




**Figure 6.8:** Ship hull mosaic rendered over different trajectories (a) Over navigation trajectory. (b) Over the estimated trajectory before the optimization. (c) Over the final optimized trajectory. (a) and (b) present blurred areas as a consequence of averaging misaligned images whereas the final mosaic shows high consistency.



**Figure 6.9:** Detail comparison of a small area in the ship hull mosaic. (a) Single frame. (b) Non-oversampled mosaic. (c) Mosaic oversampled at 3 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.



**Figure 6.10:** Qualitative comparison of the mosaics elaborated from two different datasets gathered on the same ship hull.

the whole left part of the trajectory the images are inconsistent by themselves due to the hull propeller being in motion. On the other hand, the second dataset presents constant tilt and low roll variations and is not affected by any of the aforementioned environment artifacts, yielding to a higher quality mosaic. Notice that despite this, we have presented the first dataset in order to demonstrate the ability of merging multiple transects and finding loop closures among them.

## 6.3 Marina Marciana Harbor Mapping

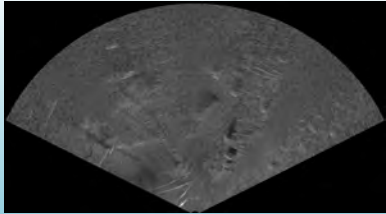
### 6.3.1 Experiment description

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), former NATO Undersea Research Centre, located in La Spezia (Italy) during which the University of Girona collaborated with CMRE. Tightly connected to ship hull inspection, the surveillance of port facilities is also key application routinely performed for security and husbandry purposes in a low visibility environment.



**Figure 6.11:** ASV equipped with the Blueview P900-130 FLS used in the Marina Marciana experiment. Image credit: CMRE.

The experiment was conducted using a Blueview P900-130 FLS mounted on CMRE's ASV (a modified vessel made by Sea Robotics, see Figure 6.11). The employed setup allows us having precise differential GPS data and heading from 2 antennas which is used as ground truth. The dataset is composed of 4416 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 a more natural environment containing typical sea floor features (e.g. vegetation, rocks) which are sparse and less prominent than those found in man-made scenarios. The acquisition sonar also has significant differences in its operating range (up to 50 m), FOV (130 deg) 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 *a priori*.

MARINA MARCIANA DATASET			
<b>Input frame</b>	<i>Size (pixels)</i>	<i>Resolution(m/pixel)</i>	<b>Example frame</b> 
	1526x848	0.057	
<b>Computation times</b>	<i>Registration (s)</i>	<i>Optimization (s)</i>	
	10316	219	
<b>Final mosaic</b>	<i>Size (m)</i>	<i>Resolution (m/pixel)</i>	
	512x352	0.057	

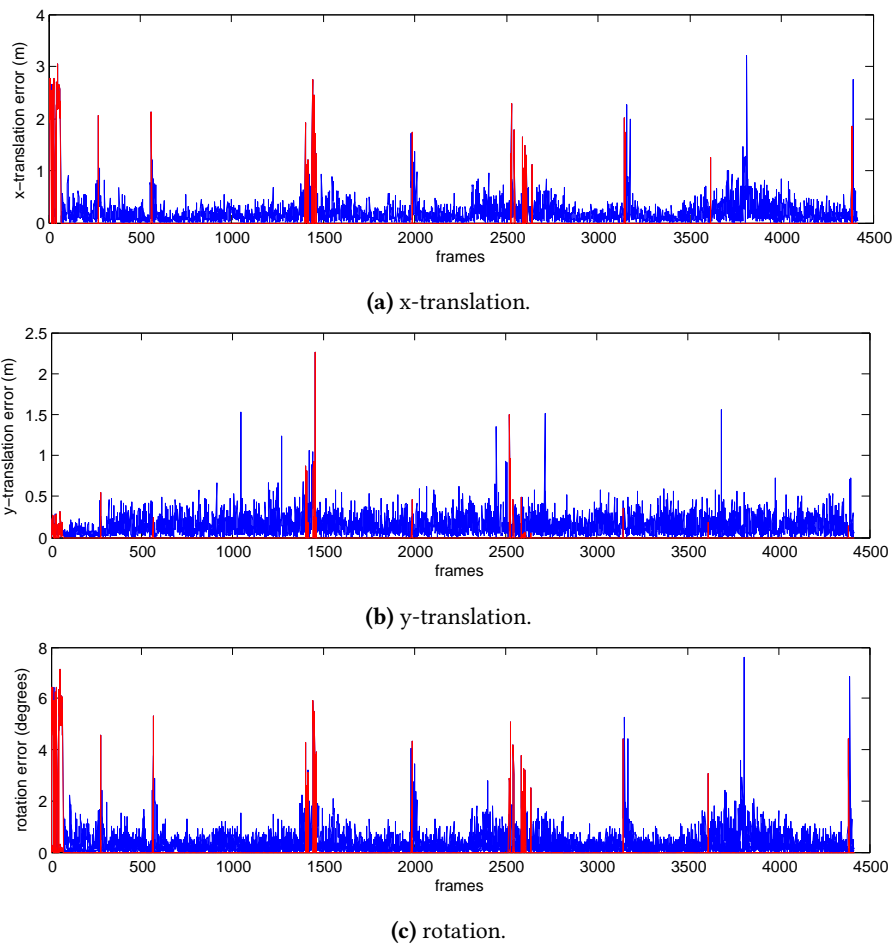
**Table 6.3:** Summary table for the Marina Marciana dataset.

### 6.3.2 Results

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 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 6.12 shows the absolute mean errors of the registration estimates for consecutive frames compared to the ground truth odometry computed from the GPS positions. The mean errors are low, being 0.23 m and 0.15 m for the  $x$  and  $y$  translations and 0.5 deg for the rotational estimates. Colored in red, we depict those consecutive registrations that have been identified as unsuccessful according to a threshold of 20 in the PSR 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 1500 and 2500. 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 highly homogeneous areas, as can be seen in the lower right side of the mosaiced area (see Figure 6.16a). In these areas the images are almost completely uniform, lacking any type of texture or intensity variation, and thus leading to a failure of the registration method which considers, with strong confidence, that there is no displacement.

Taking profit of the available ground truth, we show the error histograms of the registrations that have been identified as successful (Figure 6.13). 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 absolute GPS measurements. Note that these constraints are only

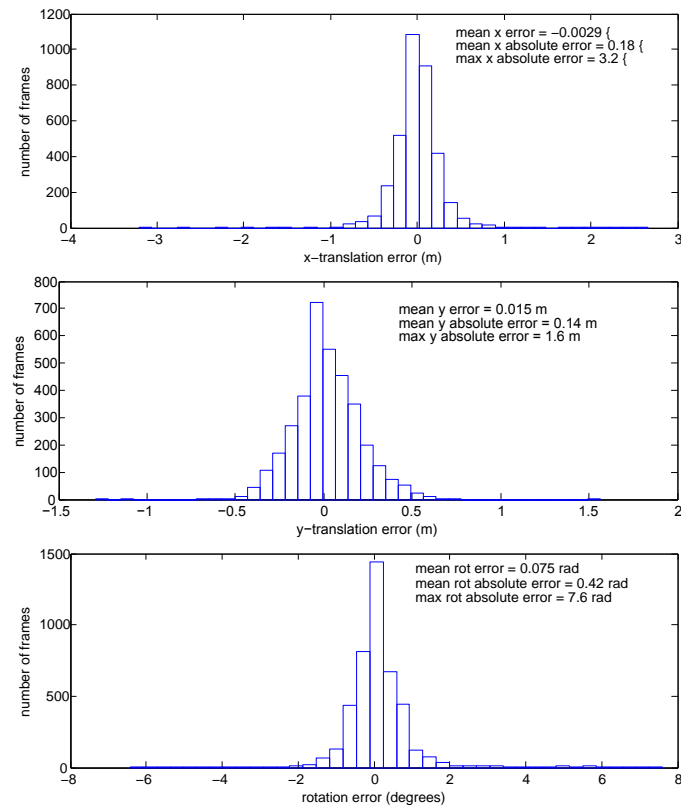


**Figure 6.12:** Absolute mean errors (in x,y and orientation) of the registration estimates for consecutive frames in the Marina Marciana dataset. Overlaid in red: errors of registrations that have been deemed as unsuccessful according to the established thresholds on the uncertainty measure.

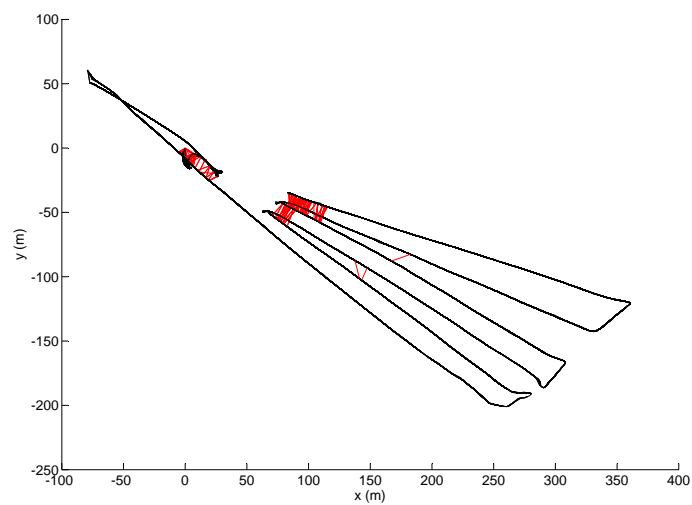
introduced in 64 of the 4415 pairs of consecutive frames. These links, together with local links from the registrations inside a window of 20 neighbouring frames, are used to estimate an initial path to determine hypothesis for temporally-far overlapping frames. In this experiment, cross-track registration 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 correctly align a small number of revisited frames crucial to enforce global consistency (Figure 6.14).

Figure 6.15 shows the GPS trajectory (in blue) 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 distributed along the trajectory, however, since the estimated trajectory

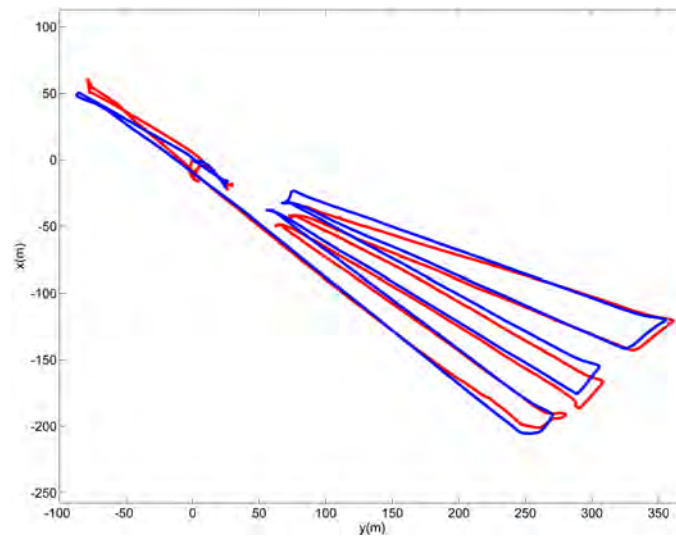




**Figure 6.13:** Error histograms for consecutive frame registrations in the Marciana Marina dataset. Only registrations considered successful under the established uncertainty threshold are taken into account. The maximum absolute errors are small and the mean of the error is around 0, indicating that 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.



**Figure 6.14:** Loop closure links detected in the Marciana Marina experiment.



**Figure 6.15:** Trajectories of the Marciana Marina experiment: GPS trajectory (in green) and estimated trajectory after the global alignment (in red).

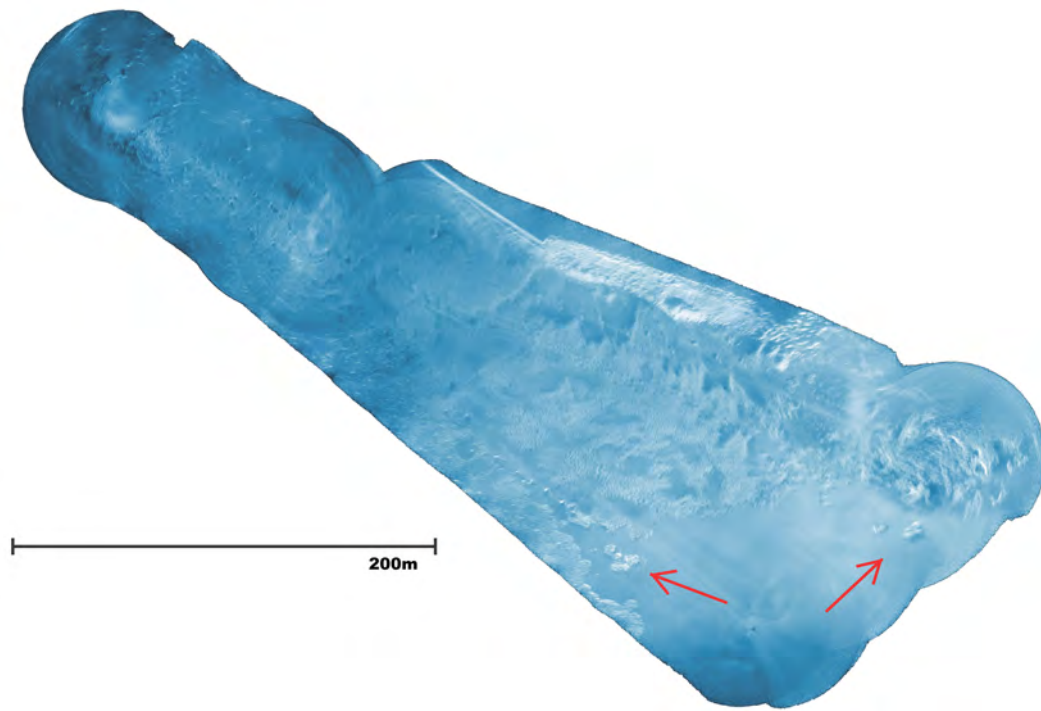
and the GPS track are fixed with respect to the first position, the error appears to be larger at the end. Notice also that the discrepancy is around 15 m, which is barely 0.7% of the total trajectory. The acoustic mosaic built using the estimated global positions (shown in Figure 6.16a) presents an 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 was gathered (Figure 6.16b) and correlate the presence of scene features (isolated rocks in the left part of the image) in both representations.

## 6.4 *Cap de Vol* Archaeological Site

### 6.4.1 Experiment description

In an effort to demonstrate the real-time capability of the mosaicing pipeline we present an experiment in the context of archaeological explorations. The results reported in this section are based on the data gathered at the *Cap de Vol* shipwreck located at Port de la Selva (Costa Brava-Spain). The experiments were conducted in July 2013 on board the THETIS archaeological ship of the Catalan Centre for Underwater Archaeology, *Centre d'Arqueologia Subaquàtica de Catalunya* (CASC) with the Girona500 AUV [Ribas et al.2012]. An ARIS Explorer 3000 sonar [Sound Metrics ARIS2013] and stereo camera system were installed in the payload area for this particular mission. The Girona500 was teleoperated over the shipwreck area at approximately



(a)

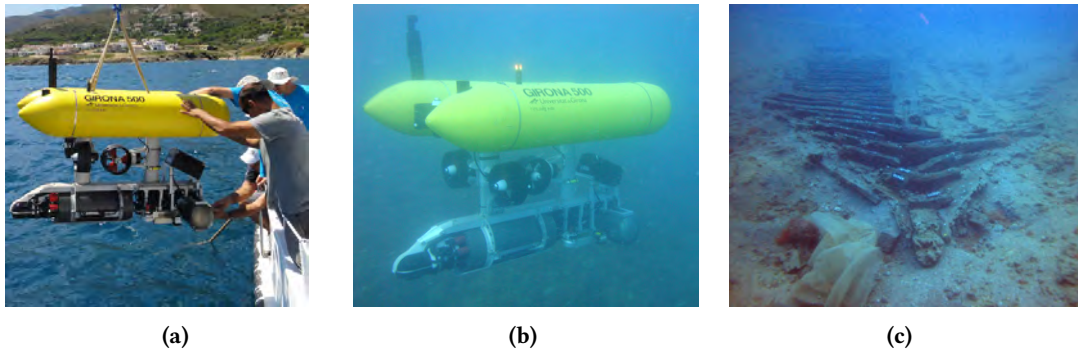


(b)

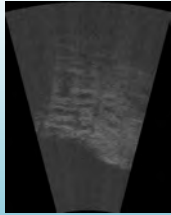
**Figure 6.16:** (a) Final mosaic of the Marciana Marina experiment (not oversampled). (b) Orthophotomap of the Marciana Marina environment. Note the presence of features that can be appreciated in both representations (pointed by red arrows).



0.2 m/s, whilst maintaining a fixed altitude (3 m) and heading to ensure more consistent shadows within the dataset. The sonar range window was set from 2.5 m to 6.5 m and the acquisition rate was set at 2 frames per second, thus ensuring a large overlap in the range direction. However, due to the lack of a pan and tilt unit and the difficulties to adjust the proper tilting angle and range without being in direct view of the scene, the images contain a large portion of blind areas that reduce the effective overlap. Several tracklines were performed to guarantee the coverage of the site as well as the possibility to establish loop closures between parallel tracks. All the original data (i.e., sonar frames, parameters, vehicle navigation) were recorded using a Robot Operative System (ROS) bag file, to be able to reproduce the experiment in real-time conditions. In total, 2720 sonar frames were recorded during the experiment which lasted approximately 22 minutes, covering an area of  $17 \times 8$  m.



**Figure 6.17:** *Cap de Vol* experiment. (a) Girona500 AUV being deployed on the site and (b) navigating over the shipwreck. (c) Detail of the *Cap de Vol* shipwreck. Image credit: CASC

CAP DE VOL DATASET			
<b>Input frame</b>	<i>Size (pixels)</i>	<i>Resolution(m/pixel)</i>	<b>Example frame</b> 
	350x450	0.008	
<b>Computation times</b>	<i>Registration</i>	<i>+ Optimization (s)</i>	
	1308		
<b>Final mosaic</b>	<i>Size (m)</i>	<i>Resolution (m/pixel)</i>	
	17x8	0.008	

**Table 6.4:** Summary table for the *Cap de Vol* dataset.

## 6.4.2 Results

We have applied the proposed mosaicing pipeline in an online fashion over the described dataset. According to the sonar frame rate, 1 frame must be processed in less than 0.5s in order to achieve real-time performance. The key point is adapting the number of registrations per frame (the more the better) so as to take advantage of all the available processing power. The Python implementation of the registration algorithm used here requires approximately 60 ms per pairwise registration,

which would result in 8 registrations in half a second. However, by using multithreading in a quad-core CPU, we are able to perform 16 registrations in 370 ms. Note that the parallelized implementation does not scale linearly with the number of CPU cores due to overhead reasons. Hence, we have set the maximum number of registrations per frame to 16 in order to leave some time for the rest of the processing.

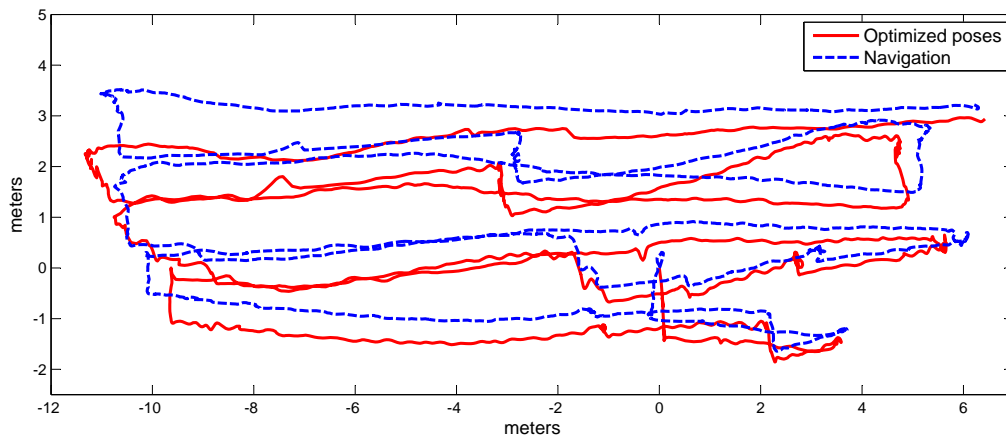
From the maximum registrations per frame, it is necessary to specify how many are devoted to registrations of consecutive frames and how many are link candidates to temporally-far frames. As described in Section 4.4.2, the decision is a trade-off between the local robustness and the possibility of finding loop-closures. In this case, we set 6 predecessor link candidates, which according to the mean speed of 0.2 m/s, ensures an overlap of over 80% and therefore high probability of obtaining correct registrations. Increasing these number of links would only reduce the possibility of finding loop closures, at the expense of obtaining more registrations with preceding frames which may not be robust. Thus, 10 out of the 16 registrations (per frame) are performed with randomly selected frames within a distance. The distance boundary to check for link hypothesis has been set to 2 meters radius from the center of the image. According to the utilized ranges, frames whose centres are farther from that distance would have less than 50% overlap and therefore are unlikely to be successfully registered.

From all computed registrations only those with a PSR higher than 18 have been included in the graph. The optimization of the graph is performed upon addition of every 10 new vertices. According to the frame rate of the experiment, this corresponds to a new optimized mosaic every 5 s. Note that within this time only the last 10 vertices contain unoptimized poses. The optimization time increases with the number of nodes, starting from less than 10 ms to a maximum of 600 ms at the end of the experiment. Due to this, some frames at the end of the sequence are skipped.

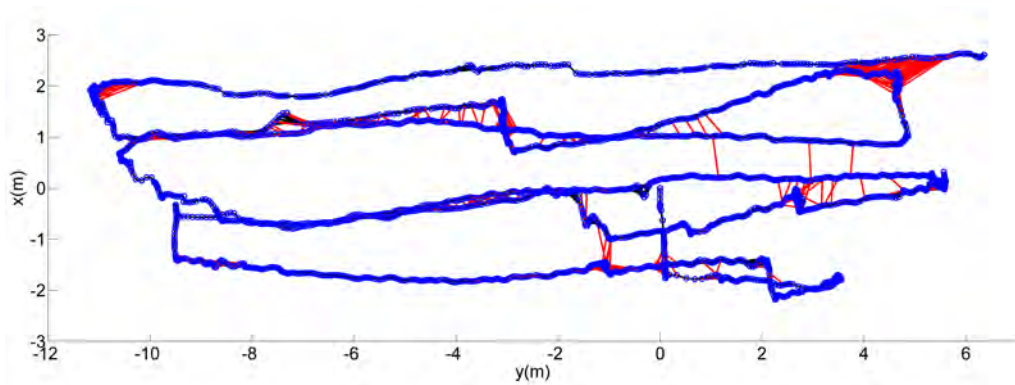
Figure 6.18 shows the navigation trajectory together with the trajectory obtained after running the online mosaicing module over the recorded data. The experiment ran successfully in real-time and concluded with 2682 vertices and a total of 13860 established links, 1235 of those corresponding to the registration of temporally far frames (see Figure 6.19).

Figure 6.20a shows the obtained final mosaic where the wooden structures of the shipwreck can be clearly seen. The rendering, also conducted in real time, is performed only by averaging the overlapping intensities at each mosaic pixel while the mosaic size is automatically expanded with the addition of new frames.

Figure 6.20b shows the mosaic obtained offline from the same experiment, involving the registration of a considerably higher number of links (71165). The offline mosaic presents slightly better definition in some areas as the higher number of registration links yields a better estimation of the final image configuration. However, the mosaic obtained in real-time is consistent and offers a good representation of the imaged scenario. An optical mosaic of the same area is provided in Fig. 6.20c for comparison purposes.

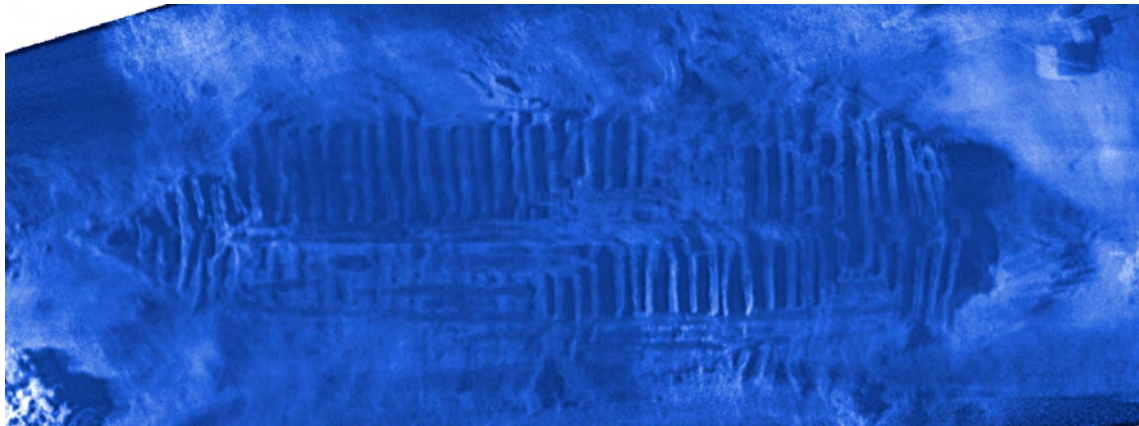


**Figure 6.18:** *Cap de Vol* experiment trajectories. In blue, the trajectory computed from the navigation data. In red, the optimized trajectory after running the online mosaicing.

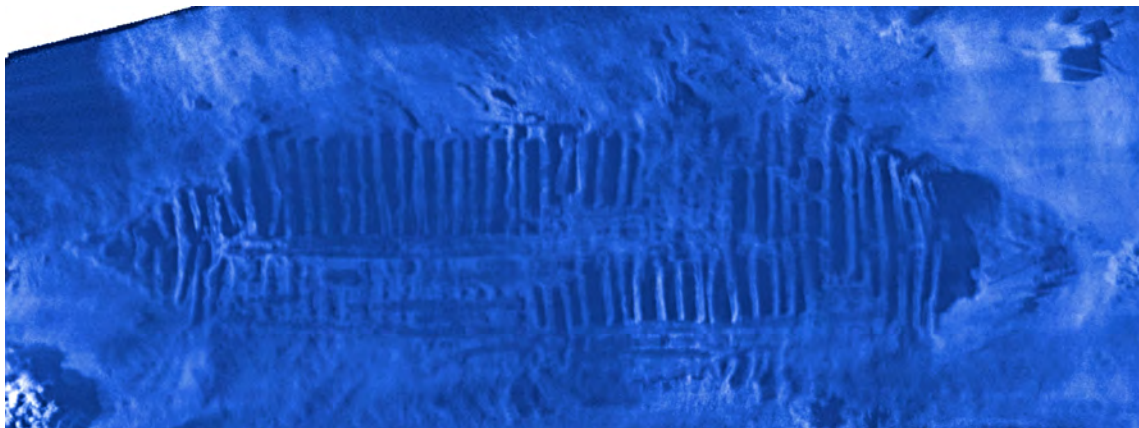


**Figure 6.19:** *Cap de Vol* experiment graph after running the online mosaicing. In blue circles, the nodes representing image poses. In black, constraints from consecutive and near-consecutive image registrations. In red, loop-closure constraints from temporally far image registrations.

The only available ground truth is a planimetric map of a portion of the wreck produced by the CASC group in 2012 by systematic measurements of the site's layout. The map has been overlaid on the mosaic in Figure 6.21. In general, the planimetry aligns well with the underlying mosaic structures. The only part where we observe a substantial deviation is in the bow, where the wooden planks do not lie flat on the seafloor and its representation gets distorted as a consequence of the 3D component. The measurements of the length from bow to stern (13 m) as well as the maximum width (4 m) are also consistent with the obtained mosaic.



(a)



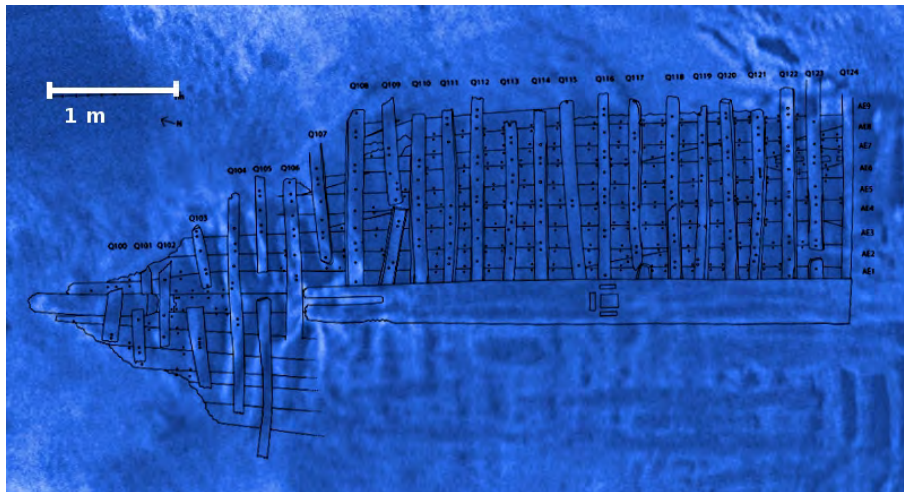
(b)



(c)

**Figure 6.20:** *Cap de Vol* shipwreck mosaic. (a) Mosaic generated on real-time (b) Mosaic generated offline. (c) Optical mosaic.





**Figure 6.21:** Portion of the *Cap de Vol* mosaic with the overlaid planimetry.

# 7

## CONCLUSIONS

---

This chapter concludes the thesis by presenting a summary of completed work in Section 7.1. We reviewing the main contributions under Section 7.2 and we discuss about the weak points of the proposed framework in Section 7.3. Finally, compelling areas for future work are outlined in Section 7.4.

## 7.1 Summary of completed work

This thesis has addressed the 2D mapping of low-visibility underwater environments by providing a full framework to generate acoustic mosaics from FLS imagery. The motivation comes from the recent breakthrough in the market of FLS devices, also termed acoustic cameras in regard to their capability to provide acoustic images at high refresh rates. The cost of this sensors is still quite high, but as their size and price decrease they become an increasingly interesting option for both ROV and AUV applications. Mapping is a quintessential application for this sort of sensors and the analogy that can be established with popular photomosaicing approaches becomes straightforward: the effective FOV of FLS images can be increased offering extended overviews of underwater areas by registering individual frames. Moreover, by using acoustic images, this can be performed regardless of the visibility conditions, though at expenses of dealing with a more challenging type of data. The current lack of effective solutions to the problem has lead to develop in this thesis a complete, versatile and efficient mosaicing pipeline tailored to the peculiarities of FLS images.

In Chapter 2, we provided some preliminary background to settle the reader into the particular characteristics of FLS imaging. We described the mode of operation of FLS devices and we discussed about the most proper imaging geometry model to be considered for the problem at hand. Taking profit of the narrow beam angle in elevation direction that typically characterizes FLS devices, we adopted an approximated orthographic model that allows us to relate two frames by a simple Euclidean transformation, while we showed that introduced errors due to this approximation remain small for typical operation ranges. We summarized as well the different challenges presented by FLS images, thus emphasizing the difficulties that are brought in to the mosaicing problem.

After that, we presented the main stages of the mosaicing pipeline. As we have seen throughout this thesis, the problem of FLS mosaicing involves a combination of different research areas including, among others, image registration, global alignment and image blending. Given the limited amount of previous efforts addressing the overall problem of FLS mosaicing, we have provided specific reviews of related research in each of the pipeline stages, highlighting the works applied on FLS imagery even though they are often in the context of other purposes rather than mosaicing.

In Chapter 3 we addressed the pairwise registration of sonar images which is a key step in the mosaicing pipeline. To overcome the instability and parameter sensitivity of feature-based registration approaches, we proposed to use a Fourier-based methodology that, by involving all image content into the registration, offers robustness to noise and the different artifacts associated with the acoustic image formation. We took profit of the inherent advantages offered by the phase correlation principle that accurately estimates translational displacements thanks to disassociating the energy content from the structure shift in the frequency domain. Building on this principle, we

adapted several aspects to cope with the multiple noise sources that can jeopardize the registration of FLS frames. We described a specific masking procedure to cope with the spectral leakage caused by the sonar footprint edges. We also presented an adaptive frequency filtering to conform to the different amounts of noise of the phase correlation matrix, thus achieving a good compromise between distinctiveness and localization of the registration peak in the spatial domain. We also studied the specific problem of rotation estimation and we proposed a simplified way of estimating it, though being only an approximation, shows better behaviour on the low SNR sonar images than other popular global rotation estimation approaches. Finally, we provided quantitative comparisons utilizing ground truth datasets to show the superior performance of the proposed registration against state-of-the-art feature-based approaches in registering both consecutive and distant FLS frames.

In Chapter 4, we have addressed the global alignment of the mosaic in order to enforce consistency between consecutive and non-consecutive image pairs. We have cast the problem as a graph optimization over the image poses, integrating spatial constraints from pairwise registrations as well as from navigation data when available. We presented a front-end to determine the constraints that should be included in the graph according to an initial estimation of the trajectory and a selection of potential overlapping candidate pairs. The workflow followed to build the graph has been provided for both offline mosaicing and the online approach, showing the interplay between the registration and global alignment modules. In the online case, constraints are added incrementally and under strict restrictions to warrant real-time operation. Besides, we proposed an heuristic to estimate the uncertainty of a pairwise registration and in this way weight appropriately the contribution of the registration constraints in the optimization. The effectiveness of the proposed heuristic in describing the uncertainty of the registration results is also demonstrated utilizing a ground truth dataset.

In Chapter 5 we tackled the rendering of the acoustic images into a smooth and informative mosaic. Taking advantage of the high frame rate of the sonar and therefore the multiple overlapping images at a given location, we adopted a blending approach based on averaging pixel intensities in order to improve the signal-to-noise ratio and resolution of the final composition with respect to the individual frames. Furthermore, we identified the different photometric irregularities that can arise from the sonar imaging configuration and we provided a set of strategies to minimize their impact both at frame and mosaic level.

In Chapter 6 we provided an extensive experimental section to validate the full proposed pipeline, showing successful results in relevant field applications such as ship-hull inspection, harbour mapping and archaeological exploration. As evidence of the framework's versatility this chapter showed the processing of datasets involving data from different sonar models, gathered through different setups and imaging diverse environments in terms of morphology and scale. Along the reported experiments we demonstrated to achieve mosaics with high self-consistency and verified accuracy when ground truth was available. Moreover, we showed examples of



mosaics obtained relying solely on FLS data and we presented an example to attest the real-time potential of the developed framework.

## 7.2 Review of main contributions

In providing an end-to-end FLS mosaicing pipeline, this thesis has advanced the current state-of-the-art for underwater mapping in low visibility conditions. We can break down this general contribution into more particular ones inside each of the framework steps:

**FLS registration** We have proposed a Fourier-based method to register a pair of FLS images in a robust and efficient way. The method, based on the phase correlation of the images, has been tailored to the particular challenges presented by FLS data. Given the difficulties of making an analytical approximation to the image noise, we have proposed a morphological approximation from the phase correlation matrix, which we use to tune a frequency filter and achieve better identification of the registration point. Furthermore, we have provided quantitative comparisons to prove its performance against other state-of-the-art feature-based techniques.

**Global alignment** We have proposed an heuristic to quantify the uncertainty of the proposed registration method and thus weight the registration constraints appropriately in the pose-based graph optimization. We have proved its effectiveness in describing the uncertainty of the registration with a quantitative evaluation on ground truth data.

**FLS mosaicing blending** We have studied the particular problem of FLS blending, identifying the different photometric irregularities that can arise and providing solutions to cope with different artifacts both at frame and mosaic level.

**Experimental evaluation** This thesis has provided experimental evaluation of the proposed method, demonstrating its applicability on real data from different sonar devices, gathered in real-world environments and in the context of relevant field applications. Moreover, we have reported successful results in different forms of processing: from offline mosaicing relying only on the FLS data, to real-time processing using sequentially incoming FLS and navigation data.

## 7.3 Framework failures

Although we have demonstrated that the proposed pipeline is able to mosaic FLS imagery with robustness and accuracy we believe we are compelled also to provide a summary of the weak points and the main issues that can lead to failures in the mosaicing framework:

**Non-properly acquired imagery** The peculiarities of FLS imaging together with the restrictions imposed by our registration algorithm must be taken into account when gathering

data to avoid the failure of the mosaicing.

The narrow beam width in the sonar elevation direction turns into critical the tilt angle of the device. If the sonar operates under a large pitch angle, not only the effective imaged area is reduced but the error introduced as a consequence of the orthographic projection model is increased. Features above the imaged surface fluctuate within a wider range of elevation angles causing larger variations between frames. The Fourier-based registration algorithm is generally robust in absorbing these variations and retrieving a registration point, however the uncertainty of the registrations increases under such circumstance and eventually, specially under an scenario with considerable relief, can lead to an inconsistent mosaic. Besides, the presence of roll motions have a similar effect, causing distortions in the image that do not conform to the adopted geometry model. Therefore data with significant roll instability is likely to end up in a fragmented and inconsistent mosaic.

A high acquisition speed can also result in the inability to generate a fully connected mosaic. If the combination of vehicle speed and sonar frame rate leads to a low overlap in consecutive images, the pose-based graph can fail to be initialized only by with image constraints. Furthermore, the number of successful registrations tends to decrease with lower overlap percentages, therefore reducing the possibilities of encountering loop closures and obtaining a consistent mosaic. Lastly, higher speeds often involve more brusque movements and rotations thus increasing vehicle instability and undesired roll motions.

Appendix A summarizes the above factors from a practical point of view, providing general guidelines to instruct a user in properly gathering FLS data.

**Completely uniform images** Despite the high tolerance of the registration method to feature-less areas, images presenting a completely uniform content can generate false positive edges, as we have seen in the Marina Marciana experiment of Section 6.3. In this cases, the registration method will retrieve, with high confidence, a zero offset that will influence the optimization even when combined with navigation edges. A possible solution to detect this situation could be to compute the entropy of the image values and discard the registration if it is lower than a predefined threshold.

**Images with highly repetitive structures** Ambiguous data association due to repetitive features in the environment can cause severe failure of least-squares based approaches. Using the proposed registration method these situations are unlikely to occur since the spatial arrangement of all image content contributes in the registration. However, notice that if this ambiguity was present it would be less critical than the issue of uniform areas. Repetitive structures will cause multiple peaks in the correlation surface, thus resulting in a low PSR that would avoid the introduction of the misleading constraint in the graph. Nonetheless, if this situation takes place in multiple occasions it would result in a lack of sufficient constraints to enforce consistency in the mosaic.

**Large rotations combined with large translations** As a consequence of the simplification that takes place in the rotation estimation method, large translations combined with large rotations are prone to generate more uncertain registrations that can be discarded. Eventually, if there are not enough links to enforce consistency that can result in a failed mosaic.

**Accumulated trajectory drift** If the initial guess at a given pose drifts too far from the truth we might not be able to generate valid link hypothesis and consequently it will be impossible to find loop closures. To avoid this situation, we must periodically maintain contact with previously visited locations. However if the revisited locations are not imaged from relatively small angle differences (less than half of the sonar's FOV) loop closures will not be found and the mosaic's trajectory will keep drifting.

## 7.4 Future work

The work developed in this thesis has laid the foundation for a promising mapping strategy in low visibility underwater environments. However, there is always room for additional improvement and further investigation. We have identified several areas that we believe are worth exploring:

**Dynamic accommodation of the imaging angle** As explained in Section 2.2.1, the sonar's tilt angle with respect to the imaged surface is critical to acquire good FLS images. When the imaged surface is not horizontal (i.e., a slope) it is necessary to adapt this angle in order to maximize the image coverage. While this can be manually adjusted using a tilt unit when operating with a ROV, it is interesting to study mechanisms to automatically accommodate this angle for AUV operations. In this way, mosaicing of more uneven terrains would be possible, obtaining a flattened 2D representation of the area despite changes in the surface gradient. Although some efforts have been devoted to estimate the underlying surface normal directly from the imagery [Negahdaripour2012a], we believe a more robust approach could be achieved by introducing an extra range sensor (e.g., a pencil beam profiler or a multibeam echosounder) in order to estimate the slope of the plane ahead of the vehicle and dynamically adapt the FLS angle using a tilt unit.

**Improvements on the optimization back-end** As mentioned in Section 4.2, we have utilized the G2O framework [Kummerle et al.2011] as an optimization back-end for the mosaicing system developed in this thesis. We have employed it as a basic tool to solve the least-squares optimization problem, concentrating mostly in providing a reliable front-end. However there is a great deal of research going on enhancements related to optimization back-ends that could improve as well the mosaic outcome, specially in terms of efficiency and robustness to outlier constraints. In this sense would be interesting to see the behaviour of approaches presented reduce or simplify the graph nodes [Johannsson et al.2013], to deal with outlier constraints [Agarwal et al.2013], to provide robustness to ambiguities [Pfungsthorn and Birk2013] or to provide a better initial guess taking into account the planarity of the tackled

problem [Carlone et al.2014]. Of special interest would be also to utilize a back-end able to deliver the uncertainty of the graph vertices. That would open the door for uncertainty driven selection of link candidates to perform registrations or, in the online case, to utilities to visualize the places that should be revisited to increase the quality of the mosaic.

**Coverage Path planning** Another area that is worth exploring is the determination of efficient paths to achieve good mapping coverage, taking into account the sonar fan shaped foot-print while trying to minimize the intensity variations to achieve more consistent mosaics.

**Opti-acoustic joint optimization** Starting with the Cap de Vol dataset presented in Section 6.4, in which we acquired both acoustic and video images, we would like to explore the possibilities of performing a joint optimization of the pose-based graph, including both camera and FLS constraints.

**Public dataset** Although FLS are becoming increasingly popular, there is not yet, to the best of the authors' knowledge, any image dataset publicly available to the research community. With the resources now available in the CIRS research group (i.e., Girona 500 AUV and the ARIS sonar), we plan to acquire and publish a dataset including AUV navigation data, FLS imagery and possibly data from other available sensors such as optical cameras. This will contribute to the improvement or new development of FLS-related techniques (e.g. image registration, navigation, SLAM, etc.), by establishing a common reference where researchers can benchmark their methods.

**Mosaicing user application** In views of developing a user-oriented application, we would like to enhance the current mosaicing pipeline with more interactive options. For instance, in an ROV case, we would like to integrate a system to guide the user in acquiring the data in proper conditions to facilitate its mosaicing (e.g., providing speed or orientation instructions to the pilot according to the feedback of the sonar and/or vehicle navigation sensors). Also, for an offline mosaicing application it would be useful to add the possibility to establish extra links by manually identifying features within an image pair.

**Acoustic visual-based image servoing** Similarly to the visual-based image servoing systems that control a vehicle using image feedback, we could envisage an scenario of target-based positioning using a FLS as a main sensor, thus allowing the operation in low-visibility underwater environments. By registering the stream of frames coming from the sonar we could obtain accurate position feedback that could be used to control the vehicle. Analogous to previous work developed at CIRS [Bechlioulis et al.2013] we could explore the use of such a system for autonomous grasping of objects from an AUV equipped with an underwater manipulator. However, the physical setup of a FLS device and a manipulator in the same vehicle complicates the feasibility of the approach, as the sonar minimum range is usually around 1m and the arm workspace can be limited. Instead, we could think of a multi-vehicle system where one vehicle is equipped with the sonar and is in charge of locating the target and communicating its position through acoustic modem commands to the vehicle that

performs the manipulation.

**Other applications** . After acquiring experience on the imaging capabilities of the FLS available at CIRS, we gained insights of what can and cannot be easily observed with this sort of sonar devices. This opens the door to explore totally different application areas that could exploit the FLS imaging. Some of these applications worth considering are, for instance, detection and measurement of flow (such as in hydrothermal vents or in underwater pipe leaks) or detection of fishing nets.



# **GUIDELINES FOR FLS IMAGE ACQUISITION**

---

The peculiarities of FLS imaging together with the restrictions imposed by our registration algorithm must be taken into account when gathering data to ensure the success of the mosaicing framework.

During the development of this thesis, we have detected that the procedure of acquiring FLS images under these constraints might not be trivial for non-experienced users. To help in these regards, we have compiled a set of guidelines and general recommendations to be followed in order to properly gather FLS data for its posterior mosaicing. While in Section 2.2 we have provided a technical introduction of the imaging geometry, here we give insights from a more practical point of view.

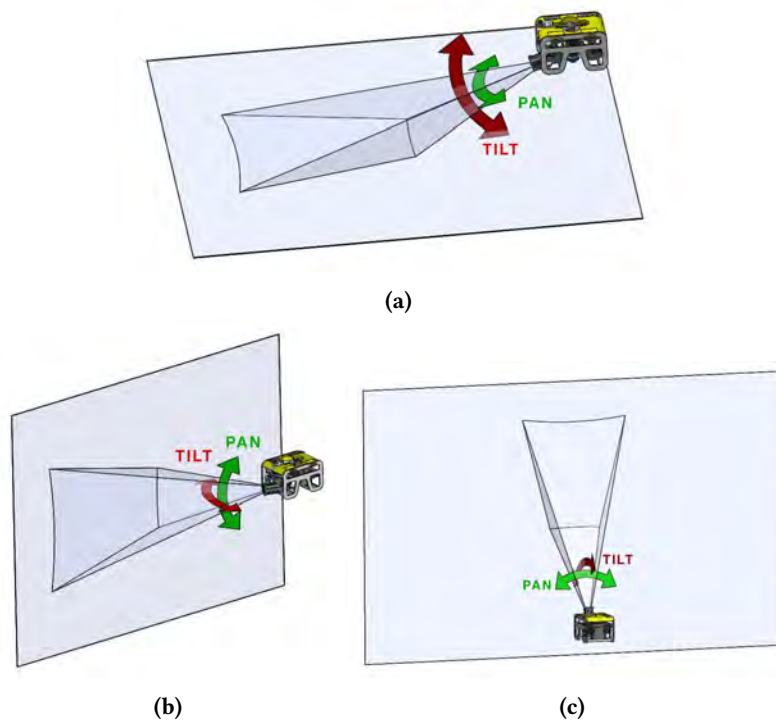
## A.1 Appropriateness of the scenario

FLS devices deliver 2D images by projecting into a plane the 3D scene points that lie within the volume limited by its horizontal (FOV) and vertical (elevation) beams. Due to this principle of operation, a FLS is not the ideal sensor to image around seamounts or abrupt terrains with strong 3D changes, but it is a valuable tool to image a fairly regular seafloor or plane surfaces such as walls. It is acceptable to have protruding objects (with some 3D component) on this base plane but the essential point is that the underlying surface should be approximately planar in order to be able to easily configure a correct imaging angle.

Notice that the requirement of a planar scenario does not imply the restriction of the plane being either horizontal or vertical. The surface to be imaged can be a slope inclined at any angle as long as we have the means to tilt the sonar appropriately to image this underlying plane as it will be explained.

## A.2 Guidelines for imaging a horizontal plane

Typical configuration scenarios are illustrated in Figure A.1. We will herein describe how to properly gather FLS images in the general case of imaging an horizontal plane (Figure A.1a). However, the guidelines can be easily extrapolated to other configurations.



**Figure A.1:** Different imaging configurations.

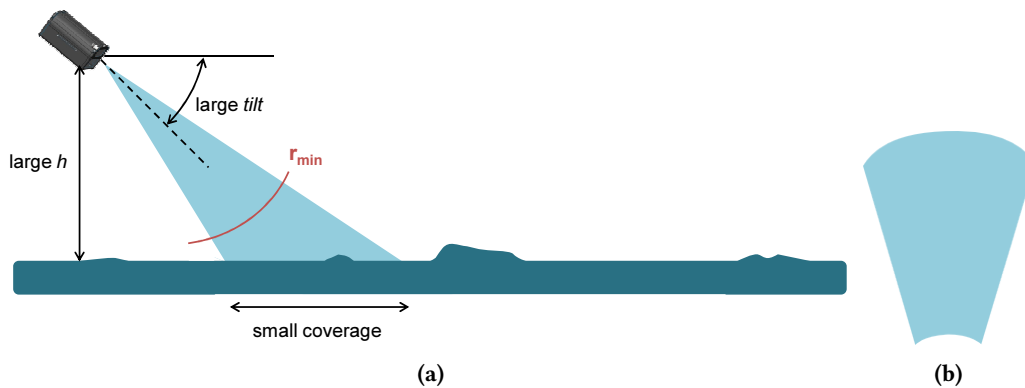


### A.2.1 Sonar imaging setup

Overall there are three main parameters to adjust: the altitude or distance to the plane, the tilt angle of the sonar with respect to the plane and the sonar ranges (minimum and maximum).

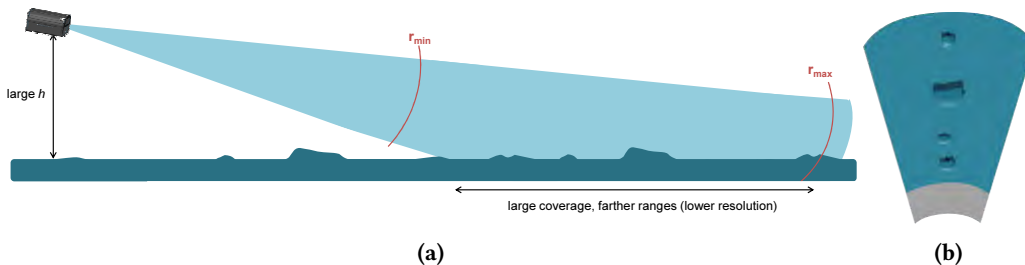
First, we will choose an altitude with respect to the base plane that is to be imaged. In general too high altitudes are a bad option as it will force us to adopt one of the following options:

- a) Tilt the sonar to a large angle, which will reduce the effective area that we cover (Figure A.2).



**Figure A.2:** Imaging with large altitude and large sonar tilt angle. (a) Setup (b) Image footprint.

- b) extend the ranges to be able to get acoustic returns from the imaged surface (Figure A.3).

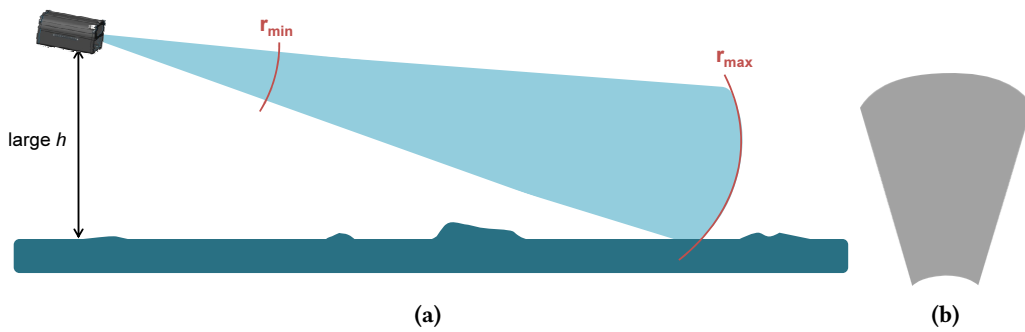


**Figure A.3:** Imaging with large altitude and large ranges. (a) Setup (b) Image footprint.

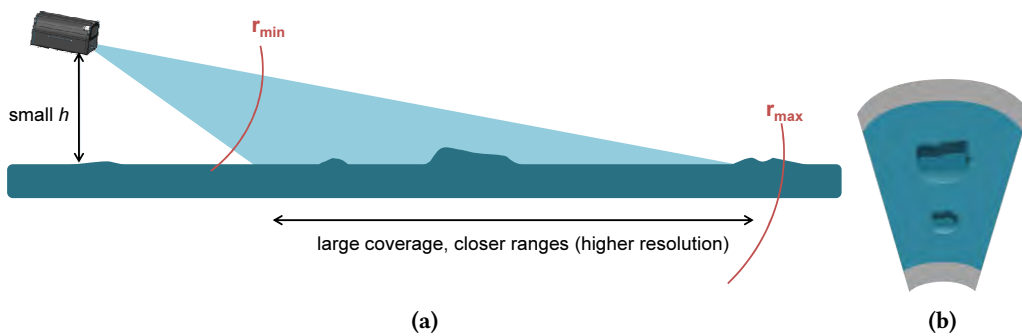
If the sonar maximum ranges are not enough we might encounter the situation of Figure A.4 in which we would obtain a full blind image.

Even in the case that the ranges can be set to longer values and reach the surface, the obtained image will be of lower resolution than imaging from a lower altitude. Therefore the advantage of imaging at a lower altitude is two fold: it allows for smaller sonar tilt angles, meaning more coverage and provides images at higher resolution due to imaging at closer ranges (Figure A.5).

Hence, in general, the lower the better, of course without sacrificing the safety of the vehicle. Around 1.5 meters is a good altitude for a sonar that has a range up to 10m. Sonars that can shoot at longer ranges will allow us to accommodate the imaging geometry easier as the altitude will not be so critical. Keeping the same altitude with respect to the plane along all the experiment



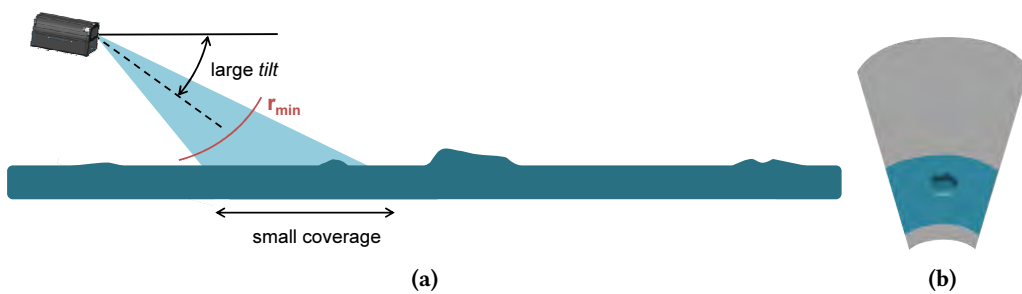
**Figure A.4:** Imaging with large altitude and insufficiently large ranges. (a) Setup (b) Image footprint.



**Figure A.5:** Imaging with low altitude. (a) Setup (b) Image footprint.

will help us to constrain the rest of the parameters that must be adjusted.

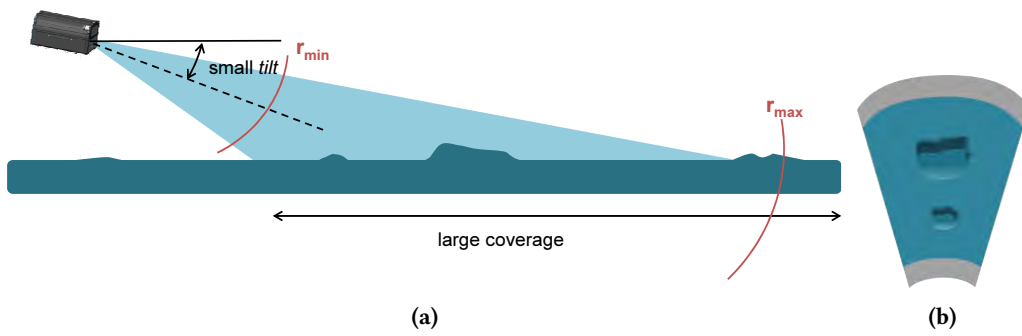
According to the set altitude two parameters remain to be adjusted. First, we will set the minimum and maximum image ranges to loose values to cover a large range window and we will start by adjusting the sonar's tilt angle. As said, a too tilted angle will reduce the effective imaged area (Figure A.6) while a small grazing angle will ensure a large coverage area (Figure A.7).



**Figure A.6:** Imaging with large tilt. (a) Setup (b) Image footprint.

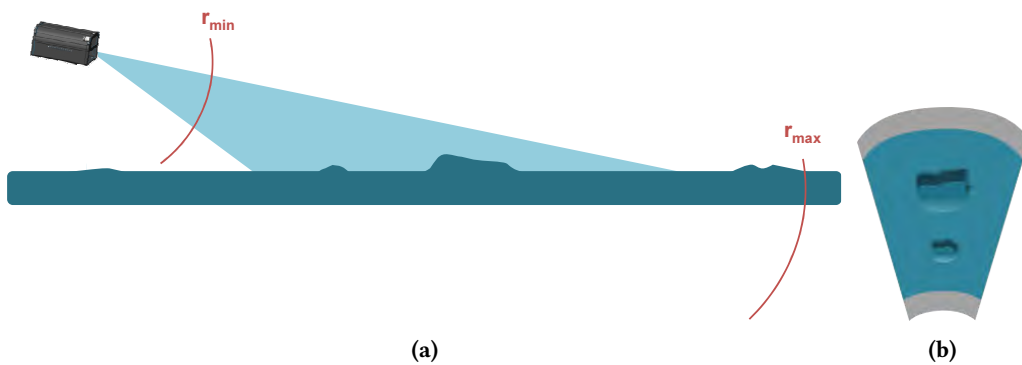
The tilt angle will also affect to the appearance of the images. A too small grazing angle would emphasize the features excessively and create elongated shadows with the risk of occluding the content at further ranges. Usually around 10-20 degrees is a good compromise. If a pan and tilt unit is available, this angle can be easily tuned.

Once the tilt is fixed, the minimum and maximum ranges should be adjusted to avoid as

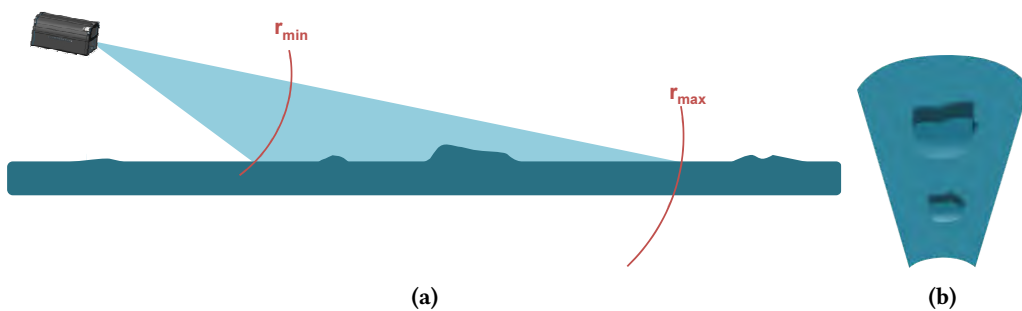


**Figure A.7:** Imaging with small tilt. (a) Setup (b) Image footprint.

much as possible blind areas (Figure A.8). Therefore we should increase the minimum range until approximately the range of the first returns and shrink the maximum range to the limit of the last returns (Figure A.9).



**Figure A.8:** Imaging with loose ranges. (a) Setup (b) Image footprint.



**Figure A.9:** Imaging with loose ranges. (a) Setup (b) Image footprint.

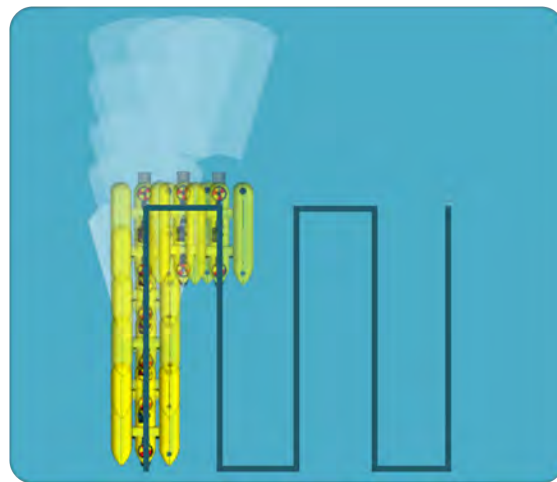
Unless required due to sudden change of the surface topology, it is advised to keep all parameters, and specially the sonar ranges, constant during all the experiment.

## A.2.2 Other considerations

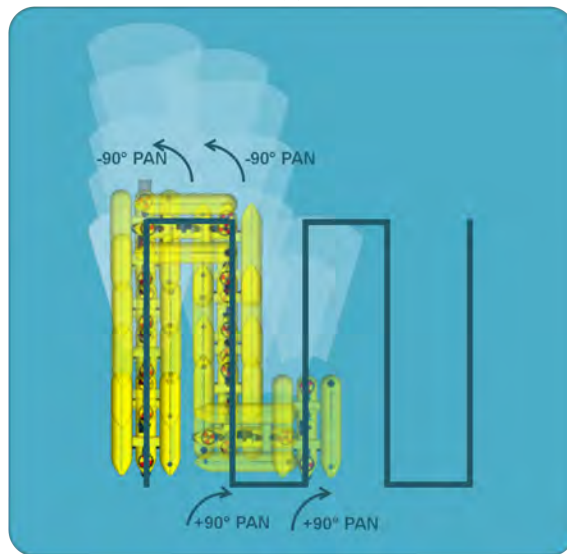
- Vehicle roll must be avoided to prevent distortions in the image content and blind areas on the image sides. The vehicle should be as stable as possible in roll and the sonar mounting as parallel as possible to the imaged plane.
- Change of the sonar's tilt (and/or the vehicle's pitch) along the experiment is not such an issue like having roll motion. It will not distort the projection of the points, however it can drastically change the portion of observed terrain within the sonar footprint. Therefore, once the sonar tilt is properly adjusted for the plane under consideration it is better to keep it constant during the acquisition. In case that, along the experiment, the base plane undergoes some general inclination change, it might be worth adjusting.
- Change of the vehicle's yaw (and/or the panning of the sonar by a pan unit) during the experiment is not a problem. However, rotating at high speeds should be avoided, specially in devices with small FOV.
- It is advised to set the sonar acquisition frame rate to the maximum value that the device allows for the configured ranges. It is as well recommended to use a slow vehicle speed, specially if the configured ranges are short and the frame rate is low.

Brusque movements should be avoided as much as possible as they can generate vehicle instability in pitch/roll DOF.

- When performing long trajectories, it is advised to revisit areas that have been previously imaged. In order to improve the consistency of the mosaic it is better to maintain roughly the same sonar orientation between different loop closures. The traditional grid survey of AUVs can be changed to conform to this restriction by making use of the sway degree of freedom to perform lateral motions instead of rotations when changing transect (Figure A.10a). Another possibility could be adjusting accordingly the orientation of the sonar after each turn with a pan and tilt unit (Figure A.10b).



(a)



(b)

**Figure A.10:** Grid surveys maintaining FLS orientation. (a) Making use of the sway DOF. (b) Adjusting the angle with a pan unit.



# B

## CHAIN DETECTION AND FOLLOWING ON FLS IMAGERY

---

In this thesis we tackled the problem with mapping with FLS. However, as introduced in Chapter 1, FLSs are a valuable tool for many applications taking place in low visibility environments. A usual practice is the detection and inspection of objects in an environment, which does not necessarily imply to build a mosaic of the area, but to be able to recognise the object in the incoming sonar frames given an *a priori* knowledge of it. This appendix encloses an example of object detection algorithm on FLS images in the context of an underwater chain inspection scenario. Tracking an underwater chain using an autonomous vehicle can be a first step towards more efficient solutions for cleaning and inspecting mooring chains. By using an FLS as a primary perception sensor we enable the vehicle operation in limited visibility conditions and overcome the turbidity arisen during marine growth removal. We developed a robust framework to perform chain following, combining perception, planning and control disciplines. We present a detection system that exploits the sonar's high frame rate and applies local pattern matching to handle the complexity of detecting link chains in acoustic images. The detection algorithm takes profit of the registration method demonstrated in this thesis to enhance the SNR of the incoming images and perform detection with a higher degree of reliability. Then, a planning system deals with the dispersed detections and determines the link waypoints that the vehicle should reach. Finally, the vehicle is guided through these waypoints using a high level controller that has been tailored to simultaneously traverse the chain and keep track of upcoming links. Furthermore, the mosaicing system presented in this thesis can be applied on the data gathered while following the chain, thus delivering an acoustic mosaic to perform a first inspection of the chain state.



## B.1 Motivation

Persistent autonomy is a subject of increasing efforts in the marine robotics community. The present work is within the context of PANDORA European FP7 project [PANDORA2014, Lane et al.2012], which aims to increase the range and complexity of underwater tasks that can be automated while reducing the need for operator supervision. To this end, one of the three core tasks of PANDORA project is to work towards a cost and time efficient solution for the cleaning and inspection of mooring chains using an AUV.

Chain moorings on floating structures such as Floating Production Storage and Offshore (FPSO) vessels are exposed to severe environmental and structural conditions. In order to avoid potential damage, chain status is monitored through periodic and exhaustive inspections. Traditional methods, which involve recovering the chain on deck or ashore, are being replaced by in situ in-water inspections using ROVs equipped with optical callipers [Hall2005] [Morandini and Legerstee2009]. However, most available solutions require prior removal of the marine biofouling so that the chain can be properly examined. Cleaning solutions range from manual brushing with divers, which is potentially hazardous and has an inherent depth limit, to high-pressure water systems deployed with ROVs [Noble Denton Europe Limited2006]. The time spent to clean strongly depends on the selected option, but in general is a tedious and slow task since the optical visibility drops drastically as the removed marine growth floats in the water. Indeed, considering the cost of ROV vessels, chain cleaning can be a significant fraction of the cost of a chain inspection program [Noble Denton Europe Limited2006].

To avoid the presence of troublesome ROV cables and reduce the cost of the deploying vessel, the PANDORA project aims to demonstrate the feasibility of using an AUV equipped with a water jet to conduct chain cleaning and inspection tasks. Our proposal is to use an AUV with a high resolution imaging sonar [Sound Metrics ARIS2013], which delivers acoustic images at near-video frame rate, in order to autonomously navigate along the chain and detect each of the links. In this way the cleaning process can be carried out regardless of the visibility conditions and the suspended marine fouling, thus speeding up the overall operation. Moreover, by producing a mosaic of the images gathered along the chain trajectory, the same methodology provides the means to perform an initial visual inspection, from which it is possible to identify some major issues or locate problematic parts that need further inspection.

Despite these advantages, the use of such a system arises several challenges that must be addressed. First, the automatic detection of the chain links in FLS images becomes a complex problem due to the inherent sonar characteristics that have been described in Section 2.3. Besides, the control of the AUV must be adapted to take into account the imaging geometry of the sonar. The vehicle location at a given instant differs from the point that is being inspected, which is located few meters ahead depending on the sonar's range configuration. Thus, to successfully follow the chain, the detection and control schemes must be tightly coupled and be able to react

in real-time. Otherwise, chain links can easily drop off the sonar's narrow FOV resulting in the vehicle losing track of the chain.

Hence, the work presented here deals with the problem of autonomously following an underwater chain as a first step towards performing chain cleaning and inspection with an AUV. An overview of the proposed framework is explained in Section B.2. Section B.3 describes the detection algorithm developed to robustly detect chain links on sonar imagery. Section B.4 covers the generation of position waypoints from the link detections. Section B.5 describes the control system of the AUV, adapted to concurrently navigate over the estimated link waypoints and keep the proper vehicle orientation to follow the upcoming links. Experiments and results are reported in Section B.6, showing the performance of the chain following framework on real experiments with a chain mock-up.

## B.2 System Overview

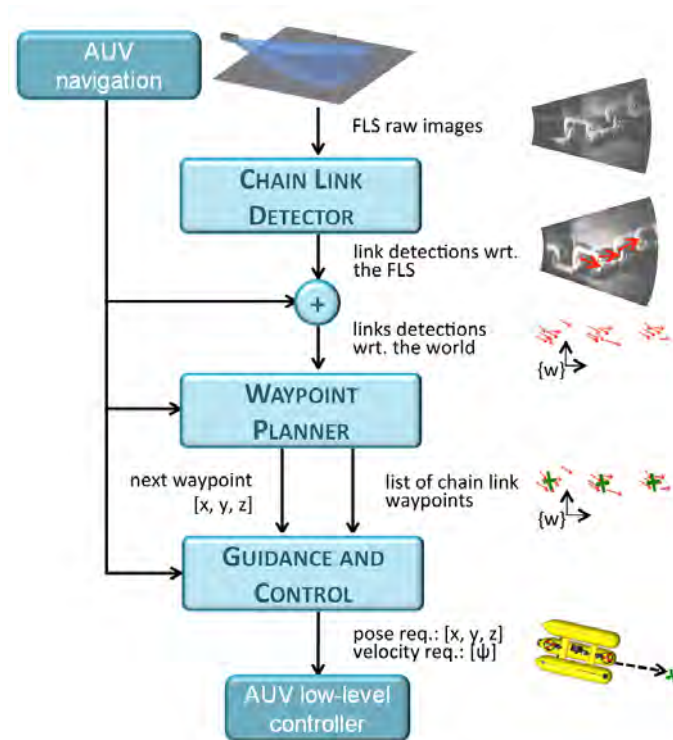
Figure B.1 outlines the developed chain following framework. Our mission scenario consists in an AUV that is deployed in the vicinity of a mooring chain end. The present methodology considers a chain lying approximately on a plane, either horizontal or vertical. Note though, that the approach could be extended to other configurations with the aid of a pan-and-tilt unit in order to set the imaging sonar to the appropriate grazing angle.

The stream of sonar images is processed by a link detection module that identifies the presence of a chain link within the sonar's FOV and delivers its position with respect to the sonar's origin. Notice that our aim is to detect each of the links as accurately as possible as the system is intended to clean them appropriately.

By taking into account the sonar position in the vehicle and the vehicle location in the world coordinate frame, link detections provided by the previous module are referenced with respect to the world coordinate frame. The different detections are grouped and associated to the corresponding chain links by a second module, named waypoint planner. This module maintains a list of all detected links and generates an ordered sequence of world waypoints that have to be visited in order to correctly follow the chain.

Finally, a third module implements a combined control system which guides the vehicle towards each sequence's waypoint while keeping the orientation to the last detected link. In this way, the vehicle visits the links that have already been identified while new ones are simultaneously detected and added to the list. This process is executed until the vehicle has reached the position of the last visible link and no link remains unvisited thus having completed the chain following procedure.

The details of the cleaning system, by means of a water jet integrated on the AUV, fall beyond the scope of the presented framework. However, a new low level controller that will be integrated



**Figure B.1:** Diagram illustrating the different steps of the chain following framework.

into the described control module is already being developed [Karras et al.2014] to be able to cope with the perturbations originated by the water jet.

### B.3 Detection of Chain Links on FLS imagery

Although the resolution of new FLS devices is progressively increasing, the inherent characteristics of the acoustic images pose a challenge to the object detection techniques that are typically used on optical images. As we have seen in Chapter 3, detection of point features in sonar images is affected by a low degree of repeatability and distinctiveness. This turns into unfeasible the detection approaches that involve matching features against a visual vocabulary of an object. A common practice in sonar object recognition is to take advantage of shadow cues. Given that chain links are not isolated objects but interlaced elements, it is difficult to exploit the use of shadows as they cannot be distinctly identified. Likewise, gradient-based or edge-based techniques become unreliable as depending on the link's position there is a wide range of different outcomes in image intensities. There are many possible intensity transitions (from link to background, link to shadow, link to link, shadow to background) together with possible sonar artefacts (e.g. cross-talk or strong reflections) that can contribute to fragment and clutter an edge map, thus complicating the task of identifying the link's contour.

The proposed detection method relies mainly on the intensities backscattered by the link itself,

which are most of the times under partial occlusions due to the imaging viewpoint or actual objects occluding parts of the chain link, such as other links or marine growth. We have approached the problem as a pattern matching using normalized cross-correlation of local templates, which allows us to detect portions of the link while others are not visible. Those detections are then robustly clustered and related according to their known dimensions and spatial location to finally identify the presence of a link and provide an estimation of its center and orientation. Fig. B.2 summarizes the link detection process, consisting of the following steps:

- **Image enhancement:** Instead of performing the detection on a raw frame, the system takes advantage of the sonar's high frame rate and registers a small number of  $n$  consecutive frames to produce an intermediate image of increased signal to noise ratio. The registration between frames is performed using Fourier-based techniques [Hurtós et al.2014] and the resulting image is formed by averaging the intensities at each point thus reducing the noise and the incidence of spurious artefacts present in the individual frames.
- **Pattern matching:** The image generated in the previous step is the basis for performing cross-correlation with local templates, each one rotated in few different orientations. These templates have been previously defined according to the morphology of the particular chain that has to be followed (e.g. for a studless chain a good template set would be composed of the four link corners and the straight side). Although many links might be present in one image, and therefore several areas can reflect a high correlation response for a given template, only the strongest location for each template is kept.
- **Clustering of detections:** To increase robustness and discard outliers, local detections are accumulated along several images, according to the displacements identified by registration. These detections are clustered by template type, ensuring that a minimum number of the same type are located within a neighbourhood in order to consider a detection as valid.
- **Link identification:** Clustered detections must be associated into groups belonging to the same link. To that end we make use of a heuristic that explores a sequence of possibilities according to the cluster's spatial location and known link dimensions. Due to link interlacing, alternate links have different inclinations with respect the main chain axis, making link elements of one side easier to be seen at a given image. Therefore the heuristic sweeps the image for the combination or sub-combination of link elements of the same side (i.e [lower-left corner/straight segment/upper-left corner] or [lower-right corner/straight segment/upper-right corner]).
- **Orientation estimation:** Once the different links have been identified, orthogonal regression is used to fit a line through the detections composing the different groups of a link. The orientation of the fitted line serves as an estimate of the link's orientation.
- **Center estimation:** Finally, using the estimated angle the location of the link's center can be estimated by projecting the known link dimensions from each detection point, assuming a 2D projection.

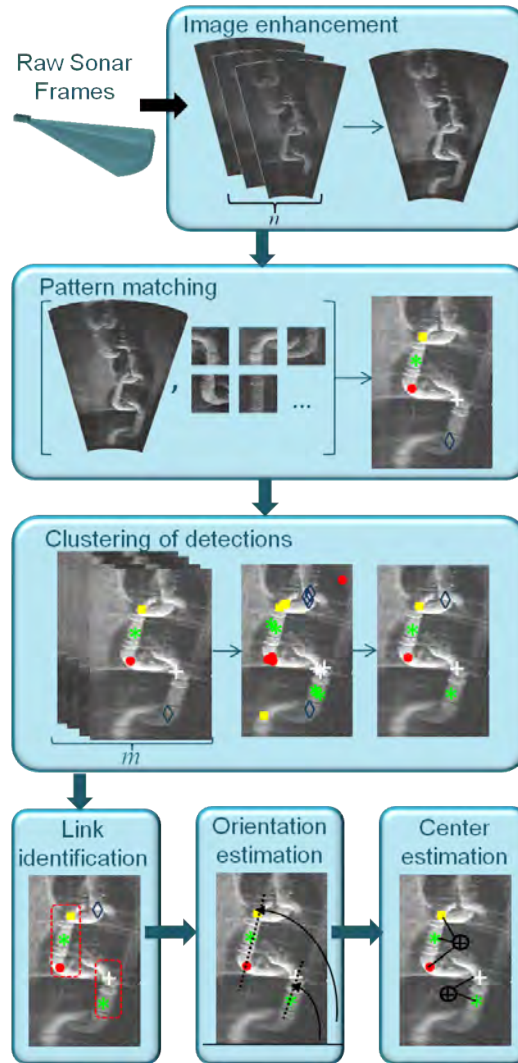


Figure B.2: Diagram of the link detection steps.

## B.4 Way-Point Planner

As the vehicle moves, the same link might be observed and detected multiple times through the module presented in previous section. However, these estimations will vary around the actual chain link center due to mainly two reasons. On the one hand, the locations found in the template matching step can slightly differ from one observation to the other as the image's intensities fluctuate due to small viewpoint changes. Therefore, the detected location of a link's part may be slightly different and so may be the final center estimation. On the other hand, since the detections are referenced in a world map by using the position of the vehicle, the accumulated navigation drift between two detections may diverge their positions. In views of that, we developed a module that: a) classifies the different estimations of the centers into clusters that belong to the same chain link and b) sorts the clusters in order to request to the vehicle the next way point that needs

to be reached to follow the chain.

Therefore, the planner module will maintain a list of ordered way points referring to the centers of the detected chain links. The output of this module at a given instant is the next waypoint that needs to be reached. Initially, the requested waypoint is the first of the list. When the vehicle reaches the desired location (considering a surrounding tolerance area) the requested waypoint will then change to the next one.

The computation of the way points is performed in three main steps. First, an  $n$ -th order polynomial curve  $y = p_n(x) = \sum_{i=0}^n a_{n-i}x^{n-i}$  is fitted on the sonar detection data  $(x_i, y_i)$ ,  $i = 1, \dots, N$ , where  $N$  is the total number of link center estimations provided by the detection module up to a given instant - see Fig. B.3a-B.3b. The polynomial fitting is performed using the least squares method. Notice that the degree  $n$  of the polynomial is a parameter that must be set according to the smoothness of the chain shape, the smoother the shape is, the lower the value of  $n$  should be selected. In general, it will be small ( $3-4^\circ$ ) as we do not expect the shape of the chain to have abrupt direction changes.

In a second step, the sonar detection data is projected onto the fitted curve, thus resulting in  $(x_i^{proj}, y_i^{proj})$ ,  $i = 1, \dots, N$  - see Fig. B.3c. The projection of each point  $(x_i, y_i)$  can be easily performed by finding the closest point to the polynomial through the solution of the following system of two equations:

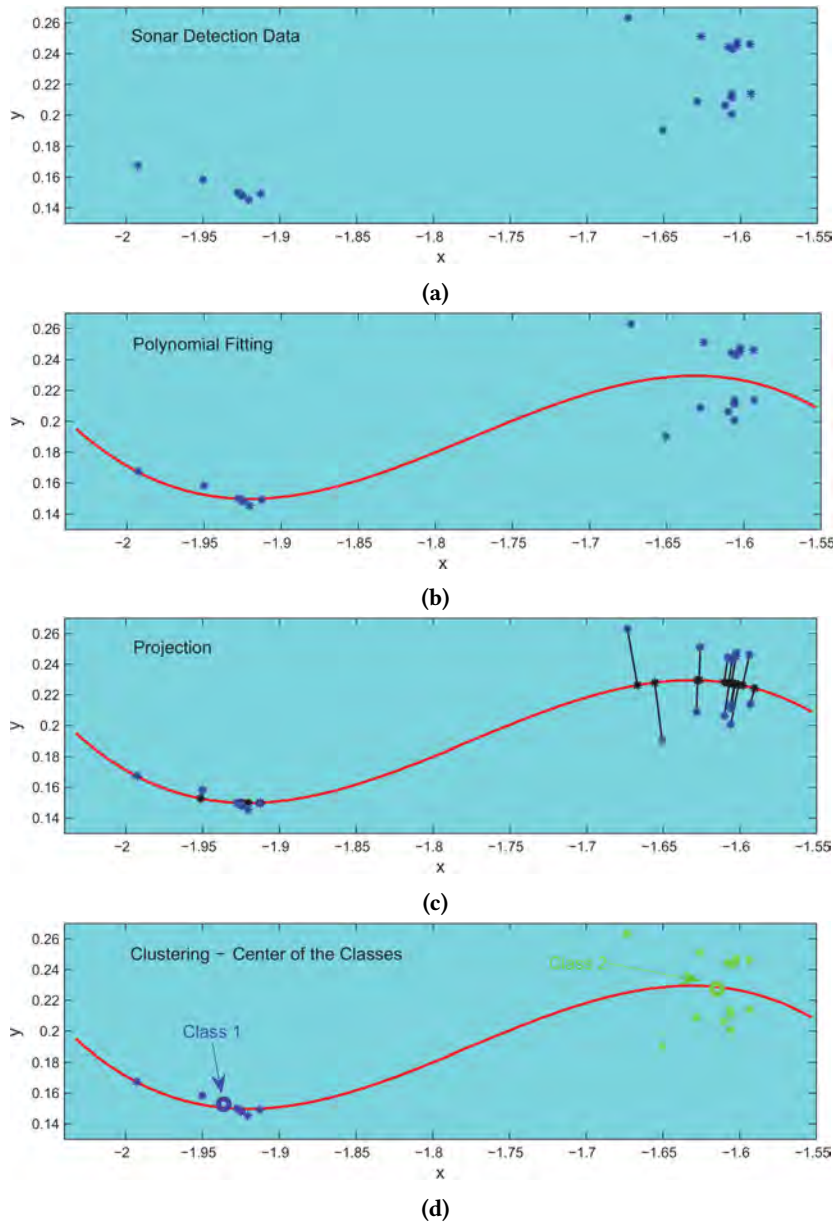
$$\begin{cases} p_n(x_i^{proj}) & = y_i^{proj} \\ (x_i - x_i^{proj}) + (y_i - y_i^{proj}) \left. \frac{dp_n(x)}{dx} \right|_{x=x_i^{proj}} & = 0. \end{cases}$$

The first equation dictates that  $(x_i^{proj}, y_i^{proj})$  belongs to the fitted polynomial  $p_n(x)$  whereas the second one imposes the fact that the vector  $[x_i - x_i^{proj}, y_i - y_i^{proj}]$  is normal to the tangent of the polynomial  $p_n(x)$  at the point  $x_i^{proj}$ .

Then, the projected data are sorted in increasing order of  $x$  values, obtaining  $(\bar{x}_i^{proj}, \bar{y}_i^{proj})$ ,  $i = 1, \dots, N$ . By computing the distances between two consecutive projected points  $\delta_i = \sqrt{(\bar{x}_{i+1}^{proj} - \bar{x}_i^{proj})^2 + (\bar{y}_{i+1}^{proj} - \bar{y}_i^{proj})^2}$ ,  $i = 1, \dots, N-1$  we find those  $j \in N_j \subset \{1, \dots, N-1\}$  for which  $\delta_j > \bar{\delta}$ , where  $\bar{\delta}$  denotes half the length of a chain link. If  $\delta_j > \bar{\delta}$ , points  $(\bar{x}_j^{proj}, \bar{y}_j^{proj})$  and  $(\bar{x}_{j+1}^{proj}, \bar{y}_{j+1}^{proj})$  are classified in different classes belonging to different chain links - see Fig. B.3d. Note that  $N_j + 1$  defines the number of total different detected chain links. Finally, we estimate the center of the detected chain links by computing the centroid of each of the projected data classes.

The estimates of the sonar detection algorithm will be updated as the AUV moves towards the desired way points; hence the described planning scheme is re-computed until there are no further updates.





**Figure B.3:** Phases of the waypoint planner computations. Blue stars: estimates from the sonar detection algorithm. Red solid line: polynomial that fits the sonar data. Black stars: projected points onto the polynomial. The two detected waypoints and their centers are depicted by stars and a circle of the same color in (d).

## B.5 Control Scheme

The control module is responsible to guide the vehicle to the world waypoints dictated by the planner. However, when the vehicle is over a waypoint, the sonar is inspecting an area located several meters ahead of the current vehicle position (depending on the configured range parameters). In order to perform a seamless chain following we need to ensure that the chain is



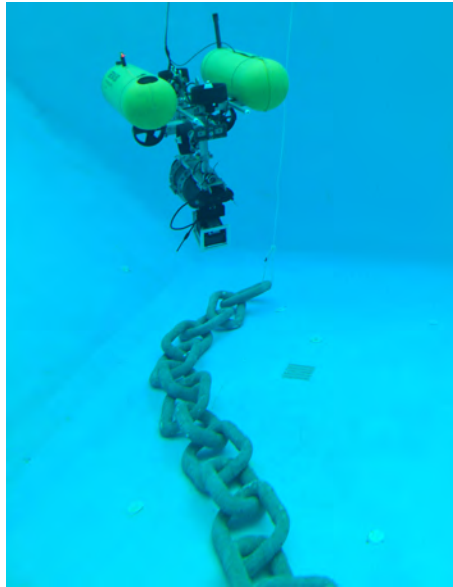
kept within the FOV of the sonar at all times, regardless of the vehicle position. To that end, our approach consists in a dual high level controller: on one side, a  $xyz$  position controller moves the vehicle to the desired waypoint. On the other, a yaw velocity controller has the purpose of keeping the farther perceivable link waypoint in the middle of the sonar's FOV. The world coordinates of the link to be centered are transformed with respect to the mid point of the sonar's FOV. In this way, the error that must be compensated in order to keep the link in the middle of the image corresponds to the  $y$  coordinate of the obtained waypoint. By feeding this error to a proportional controller we obtain the velocity setpoint that will correct the yaw rate according to the link's distance to the center. With this strategy we guarantee that the vehicle always points to the farther detected link within the FOV thus increasing the possibilities of detecting successive new ones as it moves from waypoint to waypoint.

A special situation occurs if the vehicle reaches a waypoint and there are no links inside the sonar's FOV. This circumstance can take place when: a) the vehicle is arriving to the end of the chain and the links remaining to be visited are already too close to be inside the FOV or b) the chain presents a significant deviation in its trajectory and the links drop off the FOV before being robustly detected thus losing the track of successive links. In those cases the vehicle stops and rotates 45 deg left and right in order to look around for new links, which, in case of being found are added to the planner list. The execution follows until the vehicle has reached the last waypoint in the list.

## B.6 Experiments and Results

The validation of the proposed approach has been conducted in the water tank at CIRS of the University of Girona using the Girona 500 AUV [Ribas et al.2012]. The vehicle is equipped with a complete navigation suite including a DVL, an Attitude and Heading Reference System (AHRS), and a depth sensor. For the purpose of this experiment an ARIS FLS [Sound Metrics ARIS2013] was installed in the payload area. The ARIS FLS delivers high-resolution acoustic imagery, providing an angular resolution of 0.3 deg at the expense of having a narrow FOV (30 deg). The sonar was mounted on a pan and tilt unit and the tilt angle was fixed throughout the experiments to 15 deg which is a suitable grazing angle for gathering images on the horizontal plane. A replica of a mooring chain, consisting of 13 studless links and measuring approximately 7 m, was deployed at the bottom of the tank. Figure B.4 shows a picture of the setup.

The sonar window was set to 2.5 m, extending from 1 to 3.5 m ahead of the vehicle. Such a short range was established with the aim of having a better image resolution and avoid, at the same time, reflections from the tank walls. According to the configured range samples, the acquired images (350x497 pixels) have a range resolution of 0.5 cm. The acquisition rate was set to 8 frames per second. At this frame rate, the detector module can generate an enhanced image each 3 frames and accumulate the individual template detections over 4 images thus keeping the



**Figure B.4:** Chain experiment setup at the University of Girona’s water tank: Girona 500 AUV equipped with the ARIS FLS and a chain segment deployed horizontally at the bottom.

processing in real-time and delivering link detections at 0.6 Hz.

To start the experiment the vehicle was teleoperated near one of the chain ends and was left in a position where some chain links could be seen. From then on, the described framework took control of the vehicle to drive it autonomously over the chain. We start by analysing the link detection performance. As the vehicle was moving at a low speed, many of the links were in the sonar’s FOV enough time to be detected in multiple occasions. On the course of traversing the chain the detector produced a total of 376 detections, effectively detecting 10 different chain links. In order to evaluate the accuracy of the link detector independently of potential errors introduced in the two subsequent modules, we have manually labelled the link center on those images where a link was detected. Table B.6 summarizes the number of times that each of the 13 links were detected together with the mean error computed with respect to the labelled center.

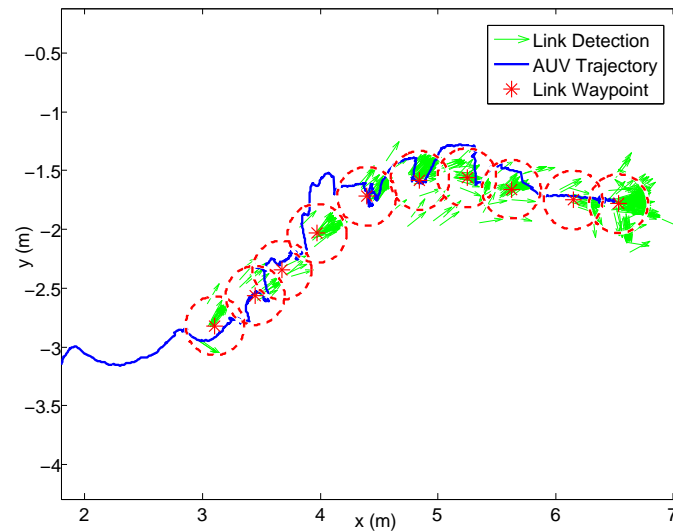
Note that the first 3 links were never detected. This is due to the limited space in the test tank that did not allow for the necessary distance between the vehicle’s starting position and the first chain links to be detected. The rest of the links were detected multiple times, depending on how much time they were in the sonar’s FOV and how well they were observed. Regarding the accuracy, all link centres have an estimated error below 15 cm. These errors are due to the variability of the template matching step as well as the assumption of the 2D projection when determining the link centres. Relatively to the link size, they are all below 1/4 of the link’s length and therefore we can consider that we achieve acceptable detection accuracy.

Link Num.	Link Detections	Mean Error (cm)	Link Num.	Link Detections	Mean Error(cm)
1	0	-	8	33	$4.5 \pm 3.9$
2	0	-	9	83	$7.9 \pm 9.5$
3	0	-	10	23	$6.8 \pm 14.1$
4	22	$12.3 \pm 3.6$	11	30	$10.2 \pm 13.3$
5	12	$6.5 \pm 9.2$	12	21	$9.9 \pm 5.9$
6	5	$7.2 \pm 5.1$	13	122	$5.6 \pm 16.3$
7	25	$4.8 \pm 4.7$			

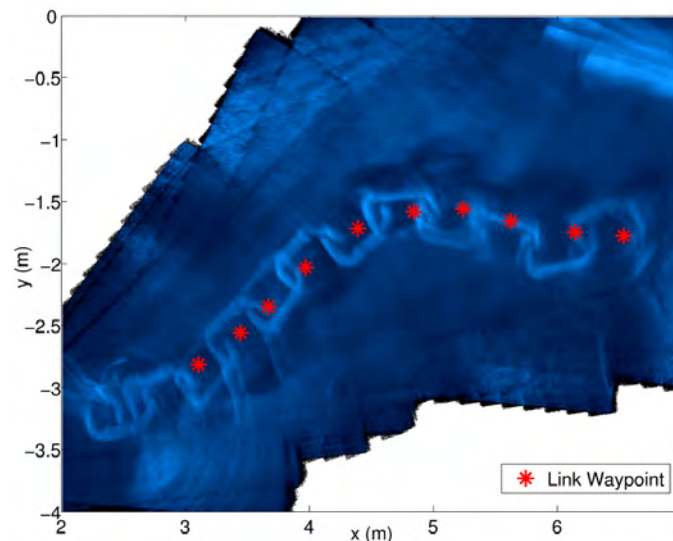
**Table B.1:** Table showing the number and accuracy (compared to manually labelled link centers) of the link detections along the chain following experiment.

Figure B.5 shows the link detections plotted on the world frame (in green), together with the waypoints (in red) that the planner module has associated to the detections and the AUV trajectory (in blue). It is important to remark that the link waypoints are dynamically recomputed during all the experiment as new detections appear by following the procedure described in Section B.4. In this sense, the depicted link waypoints are the last estimation for each link. These online adjustments in the position of the waypoints together with the small vehicle displacements induced on the turn around movements explain the slightly jagged trajectory of the AUV. Since the test chain was short, when the vehicle reached the middle of the chain the farthest links were no longer visible, thus triggering the turn around movement at each waypoint. It is also worth noting that a distance threshold of 0.03 meters was used to consider whether the vehicle had arrived to a waypoint. For this reason, the AUV trajectory does not go over the exact position of the link waypoints but effectively passes through the tolerance areas shown with red dashed circles.

Regarding the performance of the planner module, we observe, in the first place, that the links have been correctly associated to the scattered detections. Moreover, the vehicle is driven successfully through the sequence of waypoints. To verify that the performed trajectory went over the actual chain links we require the absolute positions of each link. Unfortunately this ground truth is not available. As an alternative we can compare the estimated link waypoints of the trajectory against a visual mosaic of the acquired images along the experiment. The mosaic is referenced with respect to the first sonar frame and therefore can also be referenced in world coordinates by using the vehicle location where the first frame was gathered. To build such a mosaic we have computed the pairwise registration of each frame with a window of frames in its neighborhood using Fourier-based techniques proposed in this thesis. All the obtained pose constraints have been used integrated into the pose-based graph optimization, that leads to the maximum likelihood configuration of the sonar images. Finally the frames have been fused by simply averaging the pixel intensities at each location.



**Figure B.5:** Results of the chain following experiment. AUV trajectory along the chain is indicated in blue. Green arrows represent the center and orientation of the multiple link detections. In red, the estimation of the link waypoints provided by the planner with a circle depicting the tolerance region at 0.03 m.



**Figure B.6:** Acoustic mosaic generated by registering the acquired sonar frames along the vehicle trajectory with the link waypoints overlaid in red. The link waypoints closely follow the actual link locations.

Figure B.6 shows the link waypoints overlaid on the acoustic mosaic. The obtained mosaic is, in general, visually consistent. However, the raw images were affected by several reflections and intensity artefacts causing a lack of definition in some mosaic areas. As it can be seen, the computed link waypoints closely follow the chain link centres thus testifying the accuracy of our chain following scheme. The last waypoints present small deviations, possibly due to the fact that the vehicle accumulated drift along the trajectory. To address that, in future work we could deploy absolute localization beacons or adopt an SLAM approach using the trajectory computed from the FLS registrations constraints in order to bound the navigation drift. Therefore, besides the

accuracy of the computed link waypoints, this example shows how by using the images gathered along the chain trajectory we can generate an acoustic mosaic to perform a first overall inspection of the chain state. Figure B.7 shows another example of such type of mosaic, rendered at 4 times the resolution of the original images.

Future work will involve the integration of the water jet and a new low-level controller [Karras et al.2014] to perform chain following while compensating accordingly for the force/torque disturbances produced by the jet. Besides, the chain following capability will be tested at sea where we expect a better performance as the sonar images will not be affected by the reflections and multi-paths typical of the water tank scenario.



**Figure B.7:** Example of chain mosaic.

# BIBLIOGRAPHY

---

- [AcousticView2014] AcousticView (2014). Retrieved September 1, 2014, from <http://www.acousticview.com>. (Cited on pages 7 and 8.)
- [Agarwal et al.2013] Agarwal, P., Tipaldi, G. D., Spinello, L., Stachniss, C., and Burgard, W. (2013). Robust map optimization using dynamic covariance scaling. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 62–69. IEEE. (Cited on page 108.)
- [Amat et al.1999] Amat, J., Monferrer, A., Batlle, J., and Cufi, X. (1999). Garbi: a low-cost underwater vehicle. *Microprocessors and Microsystems*, 23(2):61–67. (Cited on page 9.)
- [ANFFT Package2013] ANFFT Package (2013). Retrieved September 1, 2014, from <https://code.google.com/p/anfft>. (Cited on page 55.)
- [Armstrong et al.2006] Armstrong, R. A., Singh, H., Torres, J., Nemeth, R. S., Can, A., Roman, C., Eustice, R., Riggs, L., and Garcia-Moliner, G. (2006). Characterizing the deep insular shelf coral reef habitat of the hind bank marine conservation district (us virgin islands) using the seabed autonomous underwater vehicle. *Continental Shelf Research*, 26(2):194–205. (Cited on page 2.)
- [Aulinas et al.2010] Aulinas, J., Lladó, X., Salvi, J., and Petillot, Y. R. (2010). Feature based slam using side-scan salient objects. In *OCEANS*, pages 1–8. IEEE. (Cited on page 4.)
- [Averbuch et al.2006] Averbuch, A., Coifman, R. R., Donoho, D. L., Elad, M., and Israeli, M. (2006). Fast and accurate polar fourier transform. *Applied and Computational Harmonic Analysis*, 21(2):145–167. (Cited on page 45.)
- [Aykin and Negahdaripour2012] Aykin, M. and Negahdaripour, S. (2012). On feature extraction and region matching for forward scan sonar imaging. In *OCEANS*, pages 1–9. (Cited on pages 7, 19, 20, 22, 28, and 50.)
- [Aykin and Negahdaripour2013] Aykin, M. D. and Negahdaripour, S. (2013). On feature matching and image registration for two-dimensional forward-scan sonar imaging. *Journal of Field Robotics*, 30(4):602–623. (Cited on pages 19, 28, and 29.)
- [Balci and Foroosh2006] Balci, M. and Foroosh, H. (2006). Subpixel estimation of shifts directly in the fourier domain. *IEEE Transactions on Image Processing*, 15(7):1965–1972. (Cited on pages 31



- and 39.)
- [Batlle et al.2005] Batlle, J., Ridao, P., Garcia, R., Carreras, M., Cufi, X., El-Fakdi, A., Ribas, D., Nicosevici, T., Batlle, E., Oliver, G., et al. (2005). Uris: Underwater robotic intelligent system. *Automation for the Maritime Industries*, pages 177–203. (Cited on page 9.)
- [Baumgartner and Wales2006] Baumgartner, L. J. and Wales, N. S. (2006). *Assessment of a Dual-frequency Identification Sonar (DIDSON) for application in fish migration studies*. NSW Department of Primary Industries. (Cited on page 5.)
- [Bay et al.2006] Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf: Speeded up robust features. In *Computer Vision–ECCV 2006*, pages 404–417. Springer. (Cited on page 49.)
- [Beall et al.2011] Beall, C., Dellaert, F., Mahon, I., and Williams, S. B. (2011). Bundle adjustment in large-scale 3d reconstructions based on underwater robotic surveys. In *IEEE OCEANS Spain*, pages 1–6. IEEE. (Cited on page 59.)
- [Bechlioulis et al.2013] Bechlioulis, C. P., Karras, G. C., Nagappa, S., Palomeras, N., Kyriakopoulos, K. J., and Carreras, M. (2013). A robust visual servo control scheme with prescribed performance for an autonomous underwater vehicle. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3879–3884. IEEE. (Cited on page 109.)
- [Biber and Straßer2003] Biber, P. and Straßer, W. (2003). The normal distributions transform: A new approach to laser scan matching. In *Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, volume 3, pages 2743–2748. Ieee. (Cited on page 28.)
- [Bingham et al.2010] Bingham, B., Foley, B., Singh, H., Camilli, R., Delaporta, K., Eustice, R., Mallios, A., Mindell, D., Roman, C., and Sakellariou, D. (2010). Robotic tools for deep water archaeology: Surveying an ancient shipwreck with an autonomous underwater vehicle. *Journal of Field Robotics*, 27(6):702–717. (Cited on pages 2 and 3.)
- [Bluefin Robotics Corp.2013] Bluefin Robotics Corp. (2013). Retrieved Setember 1, 2014, from <http://www.bluefinrobotics.com>. (Cited on page 87.)
- [BlueView Technologies Inc.2013] BlueView Technologies Inc. (2013). Retrieved Setember 1, 2014, from <http://www.blueview.com>. (Cited on pages 4, 7, and 48.)
- [Bülow and Birk2011] Bülow, H. and Birk, A. (2011). Spectral registration of volume data for 6-DOF spatial transformations plus scale. In *Proceedings IEEE International Conference on Robotics and Automation (ICRA)*, pages 3078–3083. (Cited on page 30.)
- [Bülow et al.2009] Bülow, H., Birk, A., and Unnithan, V. (2009). Online generation of an underwater photo map with improved Fourier Mellin based registration. In *OCEANS - EUROPE*, pages 1 –6. (Cited on pages 29 and 43.)
- [Bülow et al.2010] Bülow, H., Pfingsthorn, M., and Birk, A. (2010). Using robust spectral registration for scan matching of sonar range data. In *7th Symposium on Intelligent Autonomous Vehicles (IAV)*. IFAC. (Cited on page 30.)

- [Burt and Adelson1983] Burt, P. J. and Adelson, E. H. (1983). A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics (TOG)*, 2(4):217–236. (Cited on page 74.)
- [Camilli et al.2010] Camilli, R., Reddy, C. M., Yoerger, D. R., Van Mooy, B. A., Jakuba, M. V., Kinsey, J. C., McIntyre, C. P., Sylva, S. P., and Maloney, J. V. (2010). Tracking hydrocarbon plume transport and biodegradation at deepwater horizon. *Science*, 330(6001):201–204. (Cited on page 2.)
- [Carlone et al.2014] Carlone, L., Aragues, R., Castellanos, J. A., and Bona, B. (2014). A fast and accurate approximation for planar pose graph optimization. *The International Journal of Robotics Research (IJRR)*, page 0278364914523689. (Cited on page 109.)
- [Carreras et al.2001] Carreras, M., Batlle, J., and Ridao, P. (2001). Hybrid coordination of reinforcement learning-based behaviors for auv control. In *Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems.*, volume 3, pages 1410–1415. IEEE. (Cited on page 9.)
- [Carreras et al.2013] Carreras, M., Candela, C., Ribas, D., Mallios, A., Magí, L., Vidal, E., Palomeras, N., and Ridao, P. (2013). Sparus ii, design of a lightweight hovering auv. In *5th International Workshop on Marine Technology (Martech)*. SARTI. (Cited on page 9.)
- [Carreras et al.2003] Carreras, M., Tiano, A., El-Fakdi, A., Zirilli, A., and Ridao, P. (2003). On the identification of non linear models of unmanned underwater vehicles. In *First IFAC workshop on guidance and control of underwater vehicles GCUVŠ03*. (Cited on page 9.)
- [Chailloux2005] Chailloux, C. (2005). Region of interest on SONAR image for non symbolic registration. In *Proceedings MTS/IEEE OCEANS*, pages 1–5. (Cited on page 30.)
- [Chen et al.2011] Chen, J., Gong, Z., Li, H., and Xie, S. (2011). A detection method based on sonar image for underwater pipeline tracker. In *Second International Conference on Mechanic Automation and Control Engineering (MACE)*, pages 3766–3769. (Cited on page 5.)
- [Chen et al.1994] Chen, Q.-S., Defrise, M., and 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. (Cited on page 43.)
- [Clarke et al.2009] Clarke, M. E., Tolimieri, N., and Singh, H. (2009). Using the seabed auv to assess populations of groundfish in untrawlable areas. In *The future of fisheries science in North America*, pages 357–372. Springer. (Cited on page 2.)
- [Cobra et al.1992] Cobra, D. T., Oppenheim, A. V., and Jaffe, J. S. (1992). Geometric distortions in side-scan sonar images: a procedure for their estimation and correction. *IEEE Journal of Oceanic Engineering*, 17(3):252–268. (Cited on page 5.)
- [Coiras et al.2007] Coiras, E., Petillot, Y., and Lane, D. M. (2007). Multiresolution 3-d reconstruction from side-scan sonar images. *IEEE Transactions on Image Processing*, 16(2):382–390. (Cited

- on page 10.)
- [Coiras et al.2004] Coiras, E., Ruiz, I. T., Petillot, Y., and Lane, D. (2004). Fusion of multiple side-scan sonar views. In *MTS/IEEE OCEANS*, volume 4, pages 2036–2054. IEEE. (Cited on page 74.)
- [Costello2008] Costello, C. (2008). Multi-reference frame image registration for rotation, translation, and scale. Technical report, DTIC Document. (Cited on page 42.)
- [De Castro and Morandi1987] De Castro, E. and Morandi, C. (1987). Registration of translated and rotated images using finite fourier transforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (5):700–703. (Cited on page 29.)
- [Dellaert and Kaess2006] Dellaert, F. and Kaess, M. (2006). Square root sam: Simultaneous localization and mapping via square root information smoothing. *The International Journal of Robotics Research (IJRR)*, 25(12):1181–1203. (Cited on pages 58 and 61.)
- [Duntley1963] Duntley, S. Q. (1963). Light in the sea. *Journal of the Optical Society of America (JOSA)*, 53(2):214–233. (Cited on page 3.)
- [Eckstein and Vlachos2009] Eckstein, A. and Vlachos, P. P. (2009). Digital particle image velocimetry (dpiv) robust phase correlation. *Measurement Science and Technology*, 20(5):055401. (Cited on page 38.)
- [El-Fakdi and Carreras2013] El-Fakdi, A. and Carreras, M. (2013). Two-step gradient-based reinforcement learning for underwater robotics behavior learning. *Robotics and Autonomous Systems*, 61(3):271–282. (Cited on page 9.)
- [Elibol2011] Elibol, A. (2011). *Efficient topology estimation for large scale optical mapping*. PhD thesis, Universitat de Girona. (Cited on page 58.)
- [Elibol et al.2011a] Elibol, A., Garcia, R., and Gracias, N. (2011a). A new global alignment approach for underwater optical mapping. *Ocean Engineering*, 38(10):1207–1219. (Cited on page 9.)
- [Elibol et al.2011b] Elibol, A., Gracias, N., Garcia, R., Gleason, A., Gintert, B., Lirman, D., and Reid, R. P. (2011b). 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, USA*, pages 50–57. (Cited on page 3.)
- [Escartín et al.2008] Escartín, J., García Campos, R., Delaunoy, O., Ferrer Vila, J., Gràcias, N. R. E., Elibol, A., Cufi i Solé, X., Neumann, L., Fornari, D., Humphris, S., et al. (2008). Globally aligned photomosaic of the lucky strike hydrothermal vent field (mid-atlantic ridge, 37° 18.5′N): Release of georeferenced data, mosaic construction, and viewing software. © *Geochemistry, Geophysics, Geosystems*, 2008, vol. 9, núm. 12, p. Q12009. (Cited on page 3.)
- [Eustice et al.2002] Eustice, R., Pizarro, O., Singh, H., and Howland, J. (2002). UWIT: underwater image toolbox for optical image processing and mosaicking in MATLAB. In *Proceedings International Underwater Technology Symposium*, pages 141–145. (Cited on page 29.)

- [Eustice et al.2008] Eustice, R. M., Pizarro, O., and Singh, H. (2008). Visually augmented navigation for autonomous underwater vehicles. *IEEE Journal of Oceanic Engineering*, 33(2):103–122. (Cited on page 2.)
- [Eustice et al.2006] Eustice, R. M., Singh, H., and Leonard, J. J. (2006). Exactly sparse delayed-state filters for view-based slam. *IEEE Transactions on Robotics*, 22(6):1100–1114. (Cited on page 58.)
- [Fairfield et al.2007] Fairfield, N., Jonak, D., Kantor, G. A., and 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. (Cited on page 4.)
- [Fan Gu and Rzhanov2006] Fan Gu, D. and Rzhanov, Y. (2006). Optimal image blending for underwater mosaics. In *OCEANS*, pages 1–5. (Cited on page 74.)
- [Ferrer et al.2007] Ferrer, J., Elibol, A., Delaunoy, O., Gracias, N., and Garcia, R. (2007). Large-area photo-mosaics using global alignment and navigation data. In *MTS/IEEE OCEANS Conference, Vancouver, Canada*, pages 1–9. (Cited on page 9.)
- [Fischler and Bolles1981] Fischler, M. and 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. (Cited on page 26.)
- [Foley et al.2009] Foley, B. P., Dellaporta, K., Sakellariou, D., Bingham, B. S., Camilli, R., Eustice, R. M., Evagelistis, D., Ferrini, V. L., Katsaros, K., Kourkoumelis, D., et al. (2009). The 2005 chios ancient shipwreck survey: new methods for underwater archaeology. *hesperia*, pages 269–305. (Cited on page 2.)
- [Foroosh et al.2002] Foroosh, H., Zerubia, J. B., and Berthod, M. (2002). Extension of phase correlation to subpixel registration. *IEEE Transactions on Image Processing*, 11(3):188–200. (Cited on pages 29, 30, 31, and 34.)
- [Galceran et al.2012] Galceran, E., Djapic, V., Carreras, M., and 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. (Cited on page 5.)
- [Galceran Yebenes2014] Galceran Yebenes, E. (2014). *Coverage path planning for autonomous underwater vehicles*. PhD thesis, Universitat de Girona. (Cited on page 9.)
- [German et al.2008] German, C. R., Yoerger, D. R., Jakuba, M., Shank, T. M., Langmuir, C. H., and Nakamura, K.-i. (2008). Hydrothermal exploration with the <i>autonomous benthic explorer</i>. *Deep Sea Research Part I: Oceanographic Research Papers*, 55(2):203–219. (Cited on page 2.)
- [Gracias et al.2009] Gracias, N., Mahoor, M., Negahdaripour, S., and Gleason, A. (2009). Fast image blending using watersheds and graph cuts. *Image and Vision Computing*, 27(5):597–607.

(Cited on page 74.)

- [Gracias et al.2013] Gracias, N., Ridao, P., Garcia, R., Escartin, J., L'Hour, M., Cibecchini, F., Campos, R., Carreras, M., Ribas, D., Palomeras, N., et al. (2013). Mapping the moon: Using a lightweight auv to survey the site of the 17th century ship 'la lune'. In *MTS/IEEE OCEANS-Bergen*, pages 1–8. IEEE. (Cited on page 10.)
- [Gracias et al.2003] Gracias, N. R., Van Der Zwaan, S., Bernardino, A., and Santos-Victor, J. (2003). Mosaic-based navigation for autonomous underwater vehicles. *IEEE Journal of Oceanic Engineering*, 28(4):609–624. (Cited on page 2.)
- [Grisetti et al.2009] Grisetti, G., Stachniss, C., and Burgard, W. (2009). Nonlinear constraint network optimization for efficient map learning. *IEEE Transactions on Intelligent Transportation Systems*, 10(3):428–439. (Cited on page 61.)
- [Hall2005] Hall, A. (2005). Cost effective mooring integrity management. In *Offshore Technology Conference*. (Cited on page 120.)
- [Harris and Stephens1988] Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50. Manchester, UK. (Cited on page 26.)
- [Hernández et al.2011] Hernández, E., Carreras, M., Antich, J., Ridao, P., and Ortiz, A. (2011). A topologically guided path planner for an auv using homotopy classes. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 2337–2343. IEEE. (Cited on page 9.)
- [Hoge2003] Hoge, W. S. (2003). A subspace identification extension to the phase correlation method [MRI application]. *IEEE Transactions on Medical Imaging*, 22(2):277–280. (Cited on pages 31 and 39.)
- [Hover et al.2012] Hover, F. S., Eustice, R. M., Kim, A., Englot, B., Johannsson, H., Kaess, M., and Leonard, J. J. (2012). Advanced perception, navigation and planning for autonomous in-water ship hull inspection. *The International Journal of Robotics Research (IJRR)*, 31(12):1445–1464. (Cited on pages 5, 28, 29, 50, 59, and 87.)
- [Hurtós et al.2012] Hurtós, N., Cufí, X., Petillot, Y., and 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)*, pages 5298–5305. (Cited on page 75.)
- [Hurtós et al.2013] Hurtós, N., Nagappa, S., Cufí, X., Petillot, Y., and Salvi, J. (2013). Evaluation of registration methods on two-dimensional forward-looking sonar imagery. In *MTS/IEEE Oceans - Bergen*. (Cited on page 28.)
- [Hurtós et al.2014] Hurtós, N., Ribas, D., Cufí, X., Petillot, Y., and Salvi, J. (2014). Fourier-based registration for robust forward-looking sonar mosaicing in low-visibility underwater environments. *Journal of Field Robotics*. (Cited on page 123.)
- [Johannsson et al.2010] Johannsson, H., Kaess, M., Englot, B., Hover, F., and Leonard, J. (2010). Imaging sonar-aided navigation for autonomous underwater harbor surveillance. In *Proceedings*

- IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4396–4403. (Cited on pages 5, 7, 17, 28, and 50.)
- [Johannsson et al.2013] Johannsson, H., Kaess, M., Fallon, M., and Leonard, J. J. (2013). Temporally scalable visual slam using a reduced pose graph. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 54–61. IEEE. (Cited on page 108.)
- [Johnson-Roberson et al.2010] Johnson-Roberson, M., Pizarro, O., Williams, S. B., and Mahon, I. (2010). Generation and visualization of large-scale three-dimensional reconstructions from underwater robotic surveys. *Journal of Field Robotics*, 27(1):21–51. (Cited on page 3.)
- [Kaess et al.2011] Kaess, M., Johannsson, H., Roberts, R., Ila, V., Leonard, J. J., and Dellaert, F. (2011). isam2: Incremental smoothing and mapping using the bayes tree. *The International Journal of Robotics Research (IJRR)*, page 0278364911430419. (Cited on page 61.)
- [Kaess et al.2008] Kaess, M., Ranganathan, A., and Dellaert, F. (2008). iSAM: Incremental smoothing and mapping. *IEEE Transactions on Robotics*, 24(6):1365–1378. (Cited on pages 58 and 61.)
- [Kaess et al.2012] Kaess, M., Williams, S., Indelman, V., Roberts, R., Leonard, J. J., and Dellaert, F. (2012). Concurrent filtering and smoothing. In *15th International Conference on Information Fusion (FUSION)*, pages 1300–1307. IEEE. (Cited on page 72.)
- [Karras et al.2014] Karras, G., Bechlioulis, C., Nagappa, S., Palomeras, N., Kyriakopoulos, K., and Carreras, M. (2014). Motion control for autonomous underwater vehicles: A robust model-free approach. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*. (Cited on pages 122 and 131.)
- [Keller et al.2005] Keller, Y., Shkolnisky, Y., and 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. (Cited on page 44.)
- [Kelley et al.2005] Kelley, D. S., Karson, J. A., Früh-Green, G. L., Yoerger, D. R., Shank, T. M., Butterfield, D. A., Hayes, J. M., Schrenk, M. O., Olson, E. J., Proskurowski, G., et al. (2005). A serpentinite-hosted ecosystem: the lost city hydrothermal field. *Science*, 307(5714):1428–1434. (Cited on page 2.)
- [Kim et al.2005] Kim, K., Neretti, N., and Intrator, N. (2005). Mosaicing of acoustic camera images. *IEE Proceedings Radar, Sonar and Navigation*, 152(4):263–270. (Cited on pages 7, 8, 23, 26, and 49.)
- [Kim et al.2006] Kim, K., Neretti, N., and Intrator, N. (2006). Video enhancement for underwater exploration using forward looking sonar. In *Advanced Concepts for Intelligent Vision Systems*, pages 554–563. Springer. (Cited on pages 26 and 58.)
- [Kim et al.2008] Kim, K., Neretti, N., and Intrator, N. (2008). Map fusion method for superresolution of images with locally varying pixel quality. *International Journal of Imaging Systems and Technology*, 18(4):242–250. (Cited on page 75.)

- [Kinsey et al.2006] Kinsey, J., Eustice, R., and 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. (Cited on page 4.)
- [Konolige et al.2010] Konolige, K., Grisetti, G., Kummerle, R., Burgard, W., Limketkai, B., and Vincent, R. (2010). Efficient sparse pose adjustment for 2d mapping. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 22–29. IEEE. (Cited on page 61.)
- [Kummerle et al.2011] Kummerle, R., Grisetti, G., Strasdat, H., Konolige, K., and Burgard, W. (2011). g2o: A general framework for graph optimization. In *Proceedings IEEE International Conference on Robotics and Automation (ICRA)*, pages 3607–3613. (Cited on pages 61 and 108.)
- [Kunz et al.2009] Kunz, C., Murphy, C., Singh, H., Pontbriand, C., Sohn, R. A., Singh, S., Sato, T., Roman, C., Nakamura, K.-i., Jakuba, M., et al. (2009). Toward extraplanetary under-ice exploration: Robotic steps in the arctic. *Journal of Field Robotics*, 26(4):411–429. (Cited on page 2.)
- [Kunz and Singh2013] Kunz, C. and Singh, H. (2013). Map building fusing acoustic and visual information using autonomous underwater vehicles. *Journal of Field Robotics*, 30(5):763–783. (Cited on page 59.)
- [Lane et al.2012] Lane, D. M., Maurelli, F., Larkworthy, T., Caldwell, D., Salvi, J., Fox, M., and Kyriakopoulos, K. (2012). Pandora: Persistent autonomy through learning, adaptation, observation and replanning. volume 12. (Cited on page 120.)
- [Leonard et al.1998] Leonard, J., Bennett, A., Smith, C., and Feder, H. (1998). Autonomous underwater vehicle navigation. Technical memorandum 98-1. *MIT Marine Robotics Laboratory*. (Cited on page 4.)
- [Li et al.2007] Li, L., Qu, Z., Zeng, Q., and Meng, F. (2007). A novel approach to image roto-translation estimation. In *IEEE International Conference on Automation and Logistics*, pages 2612–2616. (Cited on page 42.)
- [Lirman et al.2010] Lirman, D., Gracias, N., Gintert, B., Gleason, A., Deangelo, G., Dick, M., Martinez, E., and 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. (Cited on page 3.)
- [Lowe2004] Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110. (Cited on pages 26 and 49.)
- [Lu and Miliotis1997] Lu, F. and Miliotis, E. (1997). Globally consistent range scan alignment for environment mapping. *Autonomous robots*, 4(4):333–349. (Cited on page 58.)
- [Lucchese and Cortelazzo2000] Lucchese, L. and Cortelazzo, G. M. (2000). A noise-robust frequency domain technique for estimating planar roto-translations. *IEEE Transactions on Signal Processing*, 48(6):1769–1786. (Cited on pages 43 and 44.)



- [Mallios et al.2014a] Mallios, A. et al. (2014a). *Sonar scan matching for simultaneous localization and mapping in confined underwater environments*. PhD thesis, Universitat de Girona. (Cited on page 9.)
- [Mallios et al.2014b] Mallios, A., Ridao, P., Ribas, D., and Hernández, E. (2014b). Scan matching slam in underwater environments. *Autonomous Robots*, 36(3):181–198. (Cited on pages 4 and 9.)
- [Meyer and Beucher1990] Meyer, F. and Beucher, S. (1990). Morphological segmentation. *Journal of visual communication and image representation*, 1(1):21–46. (Cited on page 39.)
- [Morandini and Legerstee2009] Morandini, C. and Legerstee, F. (2009). Consistent integrity of mooring systems. *ISOPE, Osaka, Japan*. (Cited on page 120.)
- [Moré1978] Moré, J. J. (1978). The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer. (Cited on page 61.)
- [Negahdaripour2012a] Negahdaripour, S. (2012a). On 3-D scene interpretation from FS sonar imagery. In *MTS/IEEE Oceans*, pages 1–9. (Cited on pages 19, 20, and 108.)
- [Negahdaripour2012b] 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. (Cited on page 19.)
- [Negahdaripour et al.2011] Negahdaripour, S., Aykin, M. D., and Sinnarajah, S. (2011). Dynamic scene analysis and mosaicing of benthic habitats by fs sonar imaging - issues and complexities. In *Proceedings of OCEANS*, pages 1–7. (Cited on pages 7, 8, 26, and 75.)
- [Negahdaripour et al.2005] Negahdaripour, S., Firoozfam, P., and Sabzmeydani, P. (2005). On processing and registration of forward-scan acoustic video imagery. In *Proceedings 2nd Canadian Conference on Computer and Robot Vision*, pages 452–459. (Cited on pages 7, 8, 15, 23, 26, 49, and 75.)
- [Nicosevici et al.2009] Nicosevici, T., Gracias, N., Negahdaripour, S., and Garcia, R. (2009). Efficient three-dimensional scene modeling and mosaicing. *Journal of Field Robotics*, 26(10):759–788. (Cited on pages 3 and 9.)
- [Noble Denton Europe Limited2006] Noble Denton Europe Limited (2006). Research report 444 floating production system. Technical report, Health and Safety Executive (HSE). (Cited on page 120.)
- [Oceanic Imaging Consultants, Inc.2014] Oceanic Imaging Consultants, Inc. (2014). Retrieved Setember 1, 2014, from <http://www.oicinc.com>. (Cited on page 7.)
- [Paduan et al.2009] Paduan, J., Caress, D., Clague, D., Paull, C., and Thomas, H. (2009). High-resolution mapping of mass wasting, tectonic, and volcanic hazards using the mbari mapping auv. In *International Conference on seafloor mapping for geohazard assessment, Forio d’Ischia, Italy*. (Cited on page 2.)

- [Palomeras et al.2006] Palomeras, N., Carreras, M., Ridao, P., and Hernandez, E. (2006). Mission control system for dam inspection with an auv. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2551–2556. IEEE. (Cited on page 9.)
- [Palomeras et al.2012] Palomeras, N., El-Fakdi, A., Carreras, M., and Ridao, P. (2012). Cola2: A control architecture for auvs. *IEEE Journal of Oceanic Engineering*, 37(4):695–716. (Cited on page 9.)
- [PANDORA2014] PANDORA (2014). Persistent Autonomy through Learning, Adaptation, Observation and Re-Planning. Retrieved September 1, 2014, from <http://www.persistentautonomy.com>. (Cited on page 120.)
- [Pap et al.2012] Pap, L., Zou, J., and Uy, B. (2012). Accurate measurement of translational shifts by adaptively masking phase correlation. *Electronics Letters*, 48(1):20–22. (Cited on page 38.)
- [Petillot et al.2001] Petillot, Y., Tena Ruiz, I., and Lane, D. M. (2001). Underwater vehicle obstacle avoidance and path planning using a multi-beam forward looking sonar. *IEEE Journal of Oceanic Engineering*, 26(2):240–251. (Cited on page 10.)
- [Petres et al.2007] Petres, C., Pailhas, Y., Patron, P., Petillot, Y., Evans, J., and Lane, D. (2007). Path planning for autonomous underwater vehicles. *IEEE Transactions on Robotics*, 23(2):331–341. (Cited on page 10.)
- [Pfingsthorn and Birk2013] Pfingsthorn, M. and Birk, A. (2013). Simultaneous localization and mapping with multimodal probability distributions. *The International Journal of Robotics Research (IJRR)*, 32(2):143–171. (Cited on page 108.)
- [Pfingsthorn et al.2010] Pfingsthorn, M., Birk, A., Schwertfeger, S., Bülow, H., and Pathak, K. (2010). Maximum likelihood mapping with spectral image registration. In *Proceedings IEEE International Conference on Robotics and Automation (ICRA)*, pages 4282–4287. (Cited on page 63.)
- [Pizarro et al.2004] Pizarro, O., Eustice, R., and Singh, H. (2004). Large area 3d reconstructions from underwater surveys. In *MTS/IEEE OCEANS*, volume 2, pages 678–687. IEEE. (Cited on page 3.)
- [Pizer et al.1987] Pizer, S. M., Amburn, E. P., Austin, J. D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J. B., and Zuiderveld, K. (1987). Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368. (Cited on page 78.)
- [Point Cloud Library2013] Point Cloud Library (2013). Retrieved September 1, 2014, from <http://www.pointclouds.org>. (Cited on page 50.)
- [Polok et al.2013] Polok, L., Ila, V., Solony, M., Smrz, P., and Zemcik, P. (2013). Incremental block cholesky factorization for nonlinear least squares in robotics. In *Robotics: Science and Systems*, Berlin, Germany. (Cited on page 61.)

- [Porter and Duff1984] Porter, T. and Duff, T. (1984). Compositing digital images. In *ACM Siggraph Computer Graphics*, volume 18, pages 253–259. ACM. (Cited on page 74.)
- [Prados et al.2012] Prados, R., Garcia, R., Gracias, N., Escartín, J., and Neumann, L. (2012). A novel blending technique for underwater gigamosaicing. *IEEE JOURNAL OF OCEANIC ENGINEERING*, 37(4). (Cited on page 74.)
- [Prados et al.2014] Prados, R., García, R., and Neumann, L. (2014). *Image blending techniques and their application in underwater mosaicing*. Springer. (Cited on page 9.)
- [Prats et al.2012a] Prats, M., Garcia, J., Wirth, S., Ribas, D., Sanz, P., Ridao, P., Gracias, N., and Oliver, G. (2012a). Multipurpose autonomous underwater intervention: A systems integration perspective. In *20th Mediterranean Conference on Control & Automation (MED)*, pages 1379–1384. IEEE. (Cited on page 10.)
- [Prats et al.2012b] Prats, M., Ribas, D., Palomeras, N., García, J. C., Nannen, V., Wirth, S., Fernández, J. J., Beltrán, J. P., Campos, R., Ridao, P., et al. (2012b). Reconfigurable auv for intervention missions: a case study on underwater object recovery. *Intelligent Service Robotics*, 5(1):19–31. (Cited on page 9.)
- [Ranganathan et al.2007] Ranganathan, A., Kaess, M., and Dellaert, F. (2007). Fast 3d pose estimation with out-of-sequence measurements. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2486–2493. IEEE. (Cited on page 58.)
- [Reddy and Chatterji1996] Reddy, B. S. and 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. (Cited on pages 29 and 43.)
- [Reed et al.2004] Reed, S., Petillot, Y., and Bell, J. (2004). Automated approach to classification of mine-like objects in sidescan sonar using highlight and shadow information. *IEE Proceedings Radar, Sonar and Navigation*, 151(1):48–56. (Cited on page 10.)
- [Ren et al.2010] Ren, J., Jiang, J., and Vlachos, T. (2010). High-accuracy sub-pixel motion estimation from noisy images in fourier domain. *IEEE Journal in Image Processing*, 19(5):1379–1384. (Cited on page 31.)
- [Ribas et al.2007] Ribas, D., Palomeras, N., Ridao, P., Carreras, M., and Hernandez, E. (2007). Ictineuauv wins the first sauc-e competition. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 151–156. IEEE. (Cited on page 9.)
- [Ribas et al.2012] Ribas, D., Palomeras, N., Ridao, P., Carreras, M., and Mallios, A. (2012). Girona 500 AUV: From Survey to Intervention. *IEEE/ASME Transactions on Mechatronics*, 17(1):46–53. (Cited on pages 2, 9, 84, 96, and 127.)
- [Ribas et al.2008] Ribas, D., Ridao, P., Tardós, J., and Neira, J. (2008). Underwater SLAM in man made structured environments. *Journal of Field Robotics*, 25(11-12):898–921. (Cited on pages 4 and 9.)

- [Ridao et al.2002] Ridao, P., Batlle, J., and Carreras, M. (2002). O<sup>2</sup>ca<sup>2</sup>, a new object oriented control architecture for autonomy: the reactive layer. *Control Engineering Practice*, 10(8):857–873. (Cited on page 9.)
- [Ridao et al.2010] Ridao, P., Carreras, M., Ribas, D., and Garcia, R. (2010). Visual inspection of hydroelectric dams using an autonomous underwater vehicle. *Journal of Field Robotics*, 27(6):759–778. (Cited on pages 3 and 9.)
- [Roman and Singh2005] Roman, C. and Singh, H. (2005). Improved vehicle based multibeam bathymetry using sub-maps and SLAM. In *Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3662–3669, Edmonton, Canada. (Cited on page 4.)
- [Salvi et al.2008] Salvi, J., Petillot, Y., Thomas, S., and Aulinas, J. (2008). Visual slam for underwater vehicles using video velocity log and natural landmarks. In *OCEANS*, pages 1–6. IEEE. (Cited on page 2.)
- [Schmid et al.2000] Schmid, C., Mohr, R., and Bauckhage, C. (2000). Evaluation of interest point detectors. *International Journal of computer vision*, 37(2):151–172. (Cited on page 52.)
- [Schwertfeger et al.2010] Schwertfeger, S., BÅijlow, H., and 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. (Cited on page 43.)
- [Sekkati and Negahdaripour2007] Sekkati, H. and Negahdaripour, S. (2007). 3-D motion estimation for positioning from 2-D acoustic video imagery. In *Pattern Recognition and Image Analysis*, pages 80–88. Springer. (Cited on page 19.)
- [Shihavuddin et al.2013] Shihavuddin, A., Gracias, N., Garcia, R., Gleason, A. C., and Gintert, B. (2013). Image-based coral reef classification and thematic mapping. *Remote Sensing*, 5(4):1809–1841. (Cited on page 9.)
- [Singh et al.2004] Singh, H., Howland, J., and Pizarro, O. (2004). Advances in large-area photomosaicking underwater. *IEEE Journal of Oceanic Engineering*, 29(3):872–886. (Cited on page 3.)
- [Smith and Singh2006] Smith, D. and Singh, S. (2006). Approaches to multisensor data fusion in target tracking: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 18(12):1696–1710. (Cited on page 58.)
- [Smith et al.1987] Smith, R., Self, M., and Cheeseman, P. (1987). Estimating uncertain spatial relationships in robotics. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, volume 4. (Cited on page 61.)
- [Smith and Cheeseman1986] Smith, R. C. and Cheeseman, P. (1986). On the representation and estimation of spatial uncertainty. *The International Journal of Robotics Research (IJRR)*, 5(4):56–68. (Cited on page 57.)
- [Sound Metrics ARIS2013] Sound Metrics ARIS (2013). Retrieved Setember 1, 2014, from

- <http://www.soundmetrics.com/Products/ARIS-Sonars/ARIS-Explorer-3000>. (Cited on pages 22, 48, 84, 96, 120, and 127.)
- [Sound Metrics DIDSON2013] Sound Metrics DIDSON (2013). Retrieved September 1, 2014, from <http://www.soundmetrics.com/Products/DIDSON-Sonars>. (Cited on pages 7, 48, and 87.)
- [Soundmetrics Corp.2013] Soundmetrics Corp. (2013). Retrieved September 1, 2014, from <http://www.soundmetrics.com>. (Cited on page 4.)
- [Stone et al.2001] Stone, H. S., Orchard, M. T., Chang, E.-C., and Martucci, S. A. (2001). A fast direct fourier-based algorithm for subpixel registration of images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(10):2235–2243. (Cited on page 38.)
- [Stone et al.2003] Stone, H. S., Tao, B., and McGuire, M. (2003). Analysis of image registration noise due to rotationally dependent aliasing. *Journal of Visual Communication and Image Representation*, 14(2):114–135. (Cited on page 35.)
- [Tena et al.2003] Tena, I., Reed, S., Petillot, Y., Bell, J., and 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, USA. (Cited on page 4.)
- [Tena Ruiz et al.2004] Tena Ruiz, I., de Raucourt, S., Petillot, Y., and Lane, D. M. (2004). Concurrent mapping and localization using sidescan sonar. *IEEE Journal of Oceanic Engineering*, 29(2):442–456. (Cited on page 10.)
- [Tena Ruiz et al.2003] Tena Ruiz, I., Petillot, Y., and Lane, D. (2003). Improved AUV navigation using side-scan sonar. In *Proceedings OCEANS*, volume 3, pages 1261–1268 Vol.3. (Cited on page 4.)
- [Triggs et al.2000] Triggs, B., McLauchlan, P. F., Hartley, R. I., and Fitzgibbon, A. W. (2000). Bundle adjustment—A modern synthesis. In *Vision algorithms: theory and practice*, pages 298–372. Springer. (Cited on page 58.)
- [Tritech Gemini2013] Trittech Gemini (2013). Retrieved September 1, 2014, from <http://www.tritech.co.uk/product/gemini-720i-300m-multibeam-imaging-sonar>. (Cited on page 4.)
- [Tuytelaars and Mikolajczyk2008] Tuytelaars, T. and Mikolajczyk, K. (2008). Local invariant feature detectors: a survey. *Foundations and Trends® in Computer Graphics and Vision*, 3(3):177–280. (Cited on pages 26 and 49.)
- [Uyttendaele et al.2001] Uyttendaele, M., Eden, A., and Skeliski, R. (2001). Eliminating ghosting and exposure artifacts in image mosaics. In *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2. (Cited on page 74.)
- [Vaganay et al.2005] Vaganay, J., Elkins, M., Willcox, S., Hover, F., Damus, R., Dessel, S., Morash, J., and Polidoro, V. (2005). Ship hull inspection by hull-relative navigation and control. In

- Proceedings of MTS/IEEE OCEANS*, pages 761–766. (Cited on page 87.)
- [Vandrish et al.2011] Vandrish, P., Vardy, A., Walker, D., and Dobre, O. A. (2011). Side-scan sonar image registration for AUV navigation. In *Proceedings IEEE Symposium Underwater Technology (UT) and Workshop Scientific Use of Submarine Cables and Related Technologies (SSC)*, pages 1–7. (Cited on page 30.)
- [VanMiddlesworth2014] VanMiddlesworth, M. M. A. (2014). *Toward autonomous underwater mapping in partially structured 3D environments*. PhD thesis, Massachusetts Institute of Technology. (Cited on page 59.)
- [Walter2008] Walter, M. (2008). *Sparse Bayesian information filters for localization and mapping*. PhD thesis, Massachusetts Institute of Technology (MIT). (Cited on pages 5 and 15.)
- [Walter et al.2008] Walter, M., Hover, F., and Leonard, J. (2008). Slam for ship hull inspection using exactly sparse extended information filters. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 1463–1470. IEEE. (Cited on page 59.)
- [Williams et al.2010] Williams, S. B., Pizarro, O., Jakuba, M., and Barrett, N. (2010). Auv benthic habitat mapping in south eastern tasmania. In *Field and Service Robotics*, pages 275–284. Springer. (Cited on page 2.)
- [Wirth et al.2013] Wirth, S., Carrasco, P. L. N., and Codina, G. O. (2013). Visual odometry for autonomous underwater vehicles. In *MTS/IEEE OCEANS-Bergen*, pages 1–6. IEEE. (Cited on page 2.)
- [Yoerger et al.1998] Yoerger, D. R., Bradley, A. M., Walden, B. B., Singh, H., and Bachmayer, R. (1998). Surveying a subsea lava flow using the autonomous benthic explorer (abe). *International Journal of Systems Science*, 29(10):1031–1044. (Cited on page 2.)
- [Zandara et al.2013] Zandara, S., Ridao, P., Ribas, D., Mallios, A., and Palomer, A. (2013). Probabilistic surface matching for bathymetry based slam. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 40–45. IEEE. (Cited on page 9.)
- [Zhang and Negahdaripour2010] Zhang, H. and Negahdaripour, S. (2010). EKF-based recursive dual estimation of structure and motion from stereo data. *Oceanic Engineering, IEEE Journal of*, 35(2):424–437. (Cited on page 3.)
- [Zhang and Bar-Shalom2011] Zhang, S. and Bar-Shalom, Y. (2011). Optimal update with multiple out-of-sequence measurements. In *SPIE Defense, Security, and Sensing*, pages 80500F–80500F. International Society for Optics and Photonics. (Cited on page 58.)
- [Zitova and Flusser2003] Zitova, B. and Flusser, J. (2003). Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000. (Cited on pages 25 and 29.)

DECLARATION:

I hereby declare that his thesis contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference has been made.

*Girona, September 2014*

Natàlia Hurtós Vilarnau