Underwater optical and acoustic imaging: A time for fusion? A brief overview of the state-of-the-art

Fausto Ferreira, Diogo Machado, Gabriele Ferri,
Samantha Dugelay, John Potter

June 2019

Originally published in:

OCEANS 2016 MTS/IEEE Monterey, 19-23 September 2016, Monterey, CA, USA, doi: 10.1109/OCEANS.2016.7761354

About CMRE

The Centre for Maritime Research and Experimentation (CMRE) is a world-class NATO scientific research and experimentation facility located in La Spezia, Italy.

The CMRE was established by the North Atlantic Council on 1 July 2012 as part of the NATO Science & Technology Organization. The CMRE and its predecessors have served NATO for over 50 years as the SACLANT Anti-Submarine Warfare Centre, SACLANT Undersea Research Centre, NATO Undersea Research Centre (NURC) and now as part of the Science & Technology Organization.

CMRE conducts state-of-the-art scientific research and experimentation ranging from concept development to prototype demonstration in an operational environment and has produced leaders in ocean science, modelling and simulation, acoustics and other disciplines, as well as producing critical results and understanding that have been built into the operational concepts of NATO and the nations.

CMRE conducts hands-on scientific and engineering research for the direct benefit of its NATO Customers. It operates two research vessels that enable science and technology solutions to be explored and exploited at sea. The largest of these vessels, the NRV Alliance, is a global class vessel that is acoustically extremely quiet.

CMRE is a leading example of enabling nations to work more effectively and efficiently together by prioritizing national needs, focusing on research and technology challenges, both in and out of the maritime environment, through the collective Power of its world-class scientists, engineers, and specialized laboratories in collaboration with the many partners in and out of the scientific domain.

Underwater Optical and Acoustic Imaging: A Time for Fusion? A Brief Overview of the State-of-the-Art

Fausto Ferreira[1], Diogo Machado[1], Gabriele Ferri[1], Samantha Dugelay[1] and John Potter[1]

Abstract— **Underwater optical imaging has several drawbacks inherent to the physical medium such as light attenuation and turbidity. Sonars try to obviate those issues although, typically, they have lower resolutions. Combining visual and sonar data in underwater applications is still not popular, but researchers' interest in the topic is growing. However, with the advent of recent higher resolution sonar systems, the approach of combining/fusing information from both sensory modalities can bring improvements to underwater imaging. This has special interest for applications such as autonomous navigation, mapping and object recognition. In this paper we investigate the state of the art for these systems and present the most relevant approaches found in the literature.**

I. INTRODUCTION

Optical and sonar systems have been used widely in the underwater domain. Each type of system can be used in different applications and can fit different purposes thanks to its peculiar characteristics. However, for fully exploiting the features of both sensory modalities, we need to leverage the advantages offered by each modality in the different scenarios, at the same time taking into account their peculiar drawbacks which can affect their performance.

Vision based sensors have been extensively used in autonomous underwater vehicles applications. The value in optical sensors comes from their high detail which can also include colour information. Colour information is of utmost importance in underwater applications such as environmental monitoring or geological surveying e.g. coral reef monitoring [1]. High resolution data provided by optical cameras can also be important in applications such as mosaicing [2], motion estimation [3], ship hull inspection [4] and archaeologic surveys [5] among others.

Both monocular and stereo optical systems depend however on getting good features to track, which implies a scene with sufficient texture to extract them. Even in scenes with enough texture, light attenuation and water turbidity can affect the range and performance of optical sensors severely. Lighting systems can alleviate these issues. However, if the lighting is not homogeneous, it rapidly becomes itself a factor of disturbance in the optical images and high quality lighting systems are not fit (cost, power requirements) for many applications. One approach to overcome this is the use of structured light to illuminate the imaged area [6]. However, such systems are not yet compact enough to be mounted in some Autonomous Underwater Vehicles (AUVs) which limits their use. Another issue is that in underwater environments, when moving close to the seabed or performing handling tasks with a robotic arm, silt or sand is raised disturbing optical sensors performance.

Alternatively, sonars are more robust to these issues, but suffer, in general, from reduced resolution compared to optical cameras and do not provide colour. Different sonars provide different information regarding the sea bottom and sunken objects. For instance, profiling sonars such as down looking multibeams produce bathymetric representations of the bottom. Instead, synthetic aperture sonar, side-scan and forward looking sonars produce image-like data. There are however some differences with respect to optical cameras. The imaging models of these kinds of sonar imply that objects with relevant heights produce shadows. This has advantages and disadvantages. For instance, the acoustic shadow can be very different depending on the sonar view point. On the other hand, many algorithms try to recognize objects in a sonar image by looking at both the highlights and their associated shadow. At the same time, artefacts due to multi-path reflections can also occur although these can be mitigated by using the acoustic system in an appropriate configuration.

While optical cameras have limited range and need to be close to the object/environment, typically less than 20m in ideal conditions, sonar sensors can have a higher range. However, a trade-off between range and resolution has to be achieved. For instance, while some sonars such as BlueView P900 (900kHz) can reach ranges up to 100 metres or more (500 metres for Reson SeaBat at 200kHz), their resolution is insufficient to allow object recognition. Higher resolution sonars such as DIDSON from Sound Metrics (at 1.8 MHz), or more recently ARIS (also from Sound Metrics at 3 MHz), which have both been specifically developed for object identification, provide high resolution data (in the order of millimetres) almost comparable to an optical camera providing the sensor is close to the target (typically less than 10 metres).

The newer systems in the MHz range are now providing a level of detail that has allowed the implementation of innovative methods that were harder to accomplish with previous commercial systems (i.e., image mosaicing [7], navigation [4], 3D motion estimation [8] and chain inspection [9].

Recently, the first COTS high frequency 3D high resolution sonars have reached the market [10]–[12] and these types of acoustic cameras allow a direct 3D representation of the data. Obtaining instantaneous 3D representation of a scene may reduce the mission time as reconstructing 3D from 2D views

---

[1] Fausto Ferreira, Diogo Machado, Gabriele Ferri, Samantha Dugelay and John Potter are with NATO Science and Technology Organization (STO), Centre for Maritime Research and Experimentation, Viale San Bartolomeo 400, 19126 La Spezia, Italy `{firstname.surname}@cmre.nato.int`

is no longer required, potentially improving the real-time data processing and decision-making of the vehicles.

However, it must be emphasized that, even with the groundbreaking increased level of detail, sonar systems are not yet able to provide the detail provided by visual systems.

As we have outlined, both sensor modalities have intrinsic issues that diminish or increase their suitability in certain conditions. Thus, the fusion of optical and acoustic camera data is a promising approach to achieve the best of each sensor modality and to improve the overall sensing capability. However, theoretic results explaining in which conditions it is better to use one of the two modalities, or when to use a fusion of both, are still to be fully developed. Examples are sparse and aim at solving a particular problem. Nonetheless, it is useful to identify them and overview the state-of-the-art before one tackles such a task.

This paper is organized in the following way: Section II illustrates some examples of data fusion and the need to fuse acoustic and optical data as a way of counterbalancing the sensors' intrinsic issues. In Section III, we list some of the most relevant scientific work that combines optical and acoustic data in underwater environments. Section IV summarizes some of open topics that fusion brings. Finally, Section V draws some conclusions and suggests future work.

II. EXAMPLES OF IMAGERY FUSION

Data and algorithm fusion has been used in many fields and applications, both multi-sensor and intra-sensor. For instance, many applications fuse the results of different classifiers based on the same data [13]. For multi-sensor data fusion, many examples can be found in the literature both applied to robotics and other fields. There are many different kinds of fusion involving imaging sensors: laser and optics, acoustic and optics, radar and optics [14] besides more traditional data fusion schemes for navigation such as GPS and Inertial Navigation Systems (INS), INS and Doppler Velocity Log (DVL), INS and acoustic positioning systems. A good review on multisensor data fusion for autonomous underwater navigation can be found in [15].

Focusing on the fusion of acoustic and optical data, there are examples in many different application areas ranging from automatic fight detection in urban environments [16] to wearable heart rate monitors [17] or organic compound detection [18]. Other applications include weapon location [19] or medical optical-acoustic imaging [20].

In the underwater domain, the idea of combining acoustic and optical data is not new. Early attempts [21] using pencil beam sonars and laser triangulation systems established an architecture but did not present fusion results.

A. *Why fuse optical and acoustic data?*

The need to combine acoustic and optical data comes from the fact that each sensor has issues that can be counterbalanced by the fusion of data produced by the different sensory modalities. Creating efficient algorithms that explore the combination of optical and acoustic sensor modalities, harvesting the best of each, can be seen as a stepping stone in underwater environmental data collection. Using data from both vision and sonar can provide increased environmental information in scenarios where textures and/or relief are not accentuated. Both types of information can be seen as complementary in underwater environments. In this setting optic cameras have a limited operation range, constrained by light attenuation and water turbidity.

As is well-known, light attenuation limits the depth at which optical cameras can be used without artificial illumination. When artificial illumination is used, its lack of uniformity brings a new set of issues to solve in image processing. Water turbidity is another issue that cannot be avoided and whose characteristics are dependent on the typology of the environment. While some sites have clear water, other areas might have muddy waters, marine snow (a large concentration of suspended particles that scatter light and produce visual disturbances), etc. Moreover, while performing manipulation tasks in underwater environments, mud or sand can be raised increasing water turbidity. In contrast, sonars do not suffer from light attenuation and are less affected by water turbidity. They can work at higher ranges as acoustic waves penetrate turbid waters easily. The setback is that sonars typically have lower resolution than optical sensors due to the longer wavelength of sound compared to light (by a factor of at least 10^3 times larger).

The goal of this article is to briefly introduce the most relevant approaches present in the state of the art. The operational conditions that make each of the sensor modalities suitable to be used and/or their fusion are not well defined. Most studies try to solve a particular problem and are not generic enough to establish a theoretical framework that determines when to use one sensor or another or their fusion. Here the goal is to outline the state of the art. In future work, the plan is to contribute to establishing a framework.

There are many applications for novel optic-acoustic systems such as underwater archaeology, port inspection, ship hull inspection, oil & gas and in military fields.

III. STATE OF THE ART APPROACHES

Combining/fusing optical and acoustic data can be done at different levels and with different sensors. Possible simpler approaches, using echosounders and cameras such as the work presented in [22] fuse optical data collected by satellite with acoustic data collected by an echosounder for mapping coral reef habitats. The information obtained by the echosounder was used to depth correct the satellite data. Similarly, the work presented in [23] correlates in post-processing data from an echosounder and an optical camera. In both cases, there was no explicit feature matching and indeed both papers used the terms combining optical and acoustic data, not fusing.

Data fusion is more complex and involves inter-sensor fusion directly or higher level data fusion (either at feature level or even classification level). While little work has been done on inter-sensor fusion [24], some examples that focus on high level data fusion can be found in the literature and are presented in the following sub-sections.

A. Acoustic array with Optic Systems

In [25], [26], the authors presented a very interesting system that combined a single-beam depthsounder, a 2D acoustic array formed by 3 hydrophones and an optical camera. These components were rigidly connected to a frame and allowed to co-register and superimpose a 2D acoustic map for each frame of the optical video. The colour encoded acoustic frequency. The estimated source level was corrected using the depthsounder. The output was a 25 frame per second video with co-registered optical and acoustic images. This was a diver-operated system used for recording and localizing Humpback whales.

B. Pencil beam sonar with Optic Systems

Among the early attempts, several research groups used the information given by a pencil beam sonar to either improve object classification or navigation. Besides the above mentioned early work by Chantler et all [21], in [27], the authors presented a common multi-dimensional scene with data from a pencil beam sonar (mechanically scanning the area) and an optical camera for navigation purposes. In a similar fashion, [28] used a pencil beam sonar together with data from a down looking optical camera for navigation purposes. In this case, the high returns from the pencil beam (looking downwards) were used to initialize Simultaneous Localization and Mapping (SLAM) features that could be tracked in the optical camera.

C. Multibeam with Optic Systems

Other early works combined high resolution photo-mosaics with low resolution acoustic bathymetric data [29].

Later on, Sulzberger et al [30] presented a work where data from Magnetic Sensors, bottom-looking sonar and optical camera were combined in a mine hunting application. Details are lacking on the description of the fusion but the fusion was done at a classifier level meaning that only the results of each classifier were combined and no explicit feature matching took place.

Hurtos [31] was one of the few authors that tackled the problem of optic-acoustic extrinsic calibration. This work was based on fusing a multibeam sonar rigidly coupled with a camera. The calibration method was inspired by optic-laser systems extrinsic calibration techniques and was tested in a simulation environment for 3D scene reconstruction, with positive results. To gather the calibration data required for applying this technique, one should move the multimodal system in order to observe the traditional chessboard planar target at different positions and orientations in a way that allows the target to appear simultaneously in both fields of view. The author tested the system in simulation, with sensors aimed directly at a section below the vehicle, and reconstructed a 2.5D sea bottom.

Kunz [32] was able to fuse the data from a down looking multibeam and an optical camera by incorporating navigation information from these two modalities into a pose graph that estimates an AUV trajectory. Concerning the navigation problem, the two modalities are complementary in the sense that not only they can improve the pose graph estimates but they can also provide good feedback under different conditions. Furthermore, the proposed system was able to superimpose a photo-mosaic on the multibeam bathymetry.

Inglis [33] presented 3D hybrid maps also built by fusing multibeam down looking sonars with stereo optical cameras. The author, like Kunz, used pose graphs in the proposed SLAM framework. In this case, the optical modality was also used to estimate bathymetric data. A global map is divided into grid cells. Each cell was filled with data from one of the sensor modalities. To select the data modality to use in each cell, Inglis took into account several metrics like outliers or cell misalignments and filled it with the least erroneous data modality. With the described set-up, stereo optical data was generally preferred. Combining optical and acoustic data, Inglis was able to produce a more consistent bathymetric map than would have been possible using a single modality.

D. Side Scan Sonar with Optic Systems

In the work of [34], side scan data was combined with data coming from a stereo optical system. However, no explicit fusion was performed here. The optical data was used to build a 3D textured scene. The sonar data could produce a 3D bathymetric profile using a shape from shading approach. Then, each of these representations was integrated and projected as referenced layers in a multidimensional state-space map using an approach inspired in [27].

E. Forward Looking Sonar with Optic Systems

Initial work by Kalyan et al. [35] investigated the combination of optical and acoustic cameras by using an AUV with an INS, a mechanically scanned forward looking sonar and an optical camera. In this system the optical sensor was used for estimating ego-motion and to produce a 2D mosaic of the pool bottom. The forward looking sonar was merged with the inertial data and sequential scans used to build a map of the test environment. Based on the individual sensors performance, the author concluded that a future combination of both optic-acoustic sensor modalities would provide more reliable position estimation and more robust underwater navigation. However, no optical and acoustic data fusion took place.

Hover et al [36] combined data from an acoustic camera and an optical camera to improve SLAM-based navigation applied to ship-hull inspection. In this case, the features from different sensor modalities were not registered with each other, only at their own sensor level.

Another work that combines information from optical and forward looking sonar can be found in [37]. In this case, both optical and sonar data were used for target tracking purposes. The position of the targets estimated by each sensor was fused but not the images themselves. Again, no explicit fusion or extrinsic calibration was performed in this work.

Instead, the work of Negahdaripour has focused on explicitly fusing the optical and acoustic data at the feature level. This research group has derived several methods that contributed to allow us to explore the fusion of forward looking sonar and optic cameras. This work in multimodal data fusion is aimed at finding consistent methods for 3D reconstruction using high frequency forward looking sonars in underwater scenarios.

In [38], Negahdaripour derived the equations for epipolar geometry and stereo triangulation for an optic-acoustic system composed of an optic camera and a forward looking sonar. The equations were verified against both simulated optic-acoustic images of a planar grid and in an indoor pool.

Later on, the system was found to be similar in accuracy when compared to the more traditional optical stereo in short distance and clear waters [39]. However, with an increasing turbidity and at longer ranges, the accuracy of the fusion system was better than the one from the optical stereo system, which rapidly degraded.

In [40], a method for the extrinsic calibration of such systems was presented. The results with synthetic and real data validated the theoretical solution and the author suggested the addition of more views of the calibration grid to further improve the accuracy of the solution.

Negahdaripour in [41] used vision and forward looking sonar for motion estimation and target-based positioning of an underwater vehicle. Features tracked in both sensor modalities were parsed independently and were not associated. This method not only improved motion estimation when compared to methods that only make use of one of the modalities, but was also useful in overcoming the inherent ambiguities of monocular vision. This approach also allowed the 3D estimation to be carried out in broader visibility conditions.

More recently, Babaee et al. [42] proposed a method for 3D object modelling from occluding surface normals using images from correlated DIDSON and an optical camera. To solve the multi-modal registration and matching problem, the authors used the apparent contours of the objects which are more readily identifiable in both modalities. The system was aligned and configured with a negligible baseline when compared with the target distances. Using this configuration the authors were able to reconstruct objects. The proposed method allows the generation of better 3D models of objects under higher turbidity levels than could be achieved with solely optical sensor techniques.

Finally, in [43], the same authors used a similar method of occluding contours registration to estimate a 3D dense range map using a Markov Random Fields (MRF) probabilistic approach. This range map could then be used for optical image dehazing. The authors compared their approach with four other algorithms and with images under different levels of turbidity. This study is a good starting point to a more quantitative and precise study that relates the fusion to different levels of turbidity.

F. 3D Sonar Camera with Optic Systems

Another type of fusion that is gaining more attention is the fusion of 3D sonar data with 2D/3D optical data.

Early work by Fusiello et al. [44] used an optic-acoustic system for providing ROV operators with an Augmented Reality system. The proposed system used an Echoscope 1600 [10] 3D acoustic camera and a 2D optical camera in an oil rig scenario which was previously modelled in Virtual Reality Modelling Language (VRML). In this system the relative pose and position of both sensors is unknown, although it is known that the images from both sensor modalities are partially overlapping. Acoustic data was integrated with optical data after each sensing modality registered with its own model.

The structures were then segmented independently in each of the sensor modalities and registered to their model. This process was done online, and the outcome of this process was that the system continually estimated the relative poses of both sensors using a Kalman filter. Given the estimated relative poses, it was then possible to project 3D points onto the image plane, obtaining a depth image with texture information. Afterwards, this image was also converted to a depth map. Given the low contrast inherent in underwater operations, the system superimposed virtual models of objects detected with a sonar on the video feed. This system helped ROV operators to better perceive the often turbid underwater environment and ease typical underwater ROV operator tasks.

In [45], a 3D Blueview BV5000 [12] acoustic sonar was combined with 3 synchronized optical cameras forming a photogrammetric system. The fusion was done in post-processing based on matching 3D point clouds obtained by each system separately. Initial results obtained during a survey of a cave were presented in the paper.

Recently, the same group of researchers presented in [46] a new method of 3D interest point extraction and description. The goal was similar to their previous work, 3D Point cloud matching after collecting 3D data. The sensors are the same used in [45]. The authors took into account the different resolutions of each point cloud, resampling the surfaces before performing the 3D surfaces fitting. The process of combining acoustic and sonar data was semi-automatic as the user should choose the two areas that they wish to combine. This was performed also taking into account the input of the end-users (archaeologists).

Finally, in [24], the authors described an optic-acoustic 3D system with a 3D sonar (Echoscope 1600) and a stereo optical system. Both sensors give 3D representations of the scene. The authors avoided an extrinsic calibration through feature matching. Instead, the well-known Interactive Closest Point (ICP) algorithm was run to align the 3D Point clouds coming from each sensor. The calibration was performed using a chessboard, as in in the work of Negahdaripour [47] but the main difference is that there was no use of epipolar geometry to match interest points such as the corners of the chessboard. The data integration used the 3D point clouds instead as mentioned above. The system was also tested against other kinds of target in the pool and it seems promising, as the authors showed in their comparison to the state of the art.

IV. OPEN ISSUES

Among the several open research topics, we identify two as major which need to be further addressed. Specifically, the optic-acoustic extrinsic calibration and optic-acoustic feature matching.

A. Optic-Acoustic Extrinsic Calibration

To improve the performance of the data fusion, optic-acoustic systems should be extrinsically calibrated. Even if this is a common procedure for optical or pure sonar stereo systems, some of the works presented in this survey do not perform this calibration. However, having the rotation matrix and translation vector which map features from optical to sonar coordinate systems is useful to take advantage of the epipolar geometry while solving the multimodal feature association. Unfortunately, this calibration is not trivial since the method has to consider that the range of each sensor is greatly different, and that the target material and shape have to be readily recognized in both sensors' data. These are issues to deal with during the calibration process of an optic-acoustic system [24].

B. Optic-Acoustic Feature Matching

More work is needed to identify the most appropriate typologies of features and the best-suited feature matching algorithms. As in more traditional optical stereo vision, 3D reconstruction using optic-acoustic systems requires the identification of the same features in both sensor modalities. Feature matching between vision and sonar is difficult because of the different imaging models and resolutions for each sensor that represent the same feature differently. Generally there is no correspondence between optic (textures) and acoustic (ranges) features. Some authors [43] solve the matching problem by using structural features like contours and edges that are available in both sensor modalities. Others use ICP algorithms to combine 3D Point clouds [24] or 3D feature maching [46]. Other works superimpose the data without registering explicitly and taking advantage of the georeferencing [22], [34] of the data. More work is needed on choosing the best kind of features and type of feature matching.

V. CONCLUSION

This paper has provided a brief overview of recent developments in optic-acoustic systems. These systems arose as a way of obviating intrinsic problems inherent with each sensor modality. By combining optical and acoustic systems, the advantages of both sensory modalities can be exploited. Furthermore, their complementarity can lead to improved performance in several scenarios. This development became possible mainly due to the technological advances in high resolution sonars, both profiling and imaging. Research interest in the topic is growing and several groups are working in this emergent field. Multibeam down looking sonars, forward looking sonars and 3D sonars can be fused with optical cameras with different goals. Many works propose to combine 2D or 3D data in post-processing by simply superimposing the data or registering 3D point clouds. Other works fuse bathymetric data with optical data for navigation purposes. However, few works have tackled the complex problem of calibrating an optic-acoustic system and the associated feature matching. We envision that achieving a correct calibration of the optic-acoustic system would be highly beneficial in terms of achievable performance. This paper selects examples from the state-of-the-art demonstrating the broad spectrum of applications and approaches. The research in this area is becoming more mature but more effort is needed to make this field popular among the underwater community. The recent high-resolution imaging sonars (including 3D sonars) open future research avenues that should develop this field. At the same time, theoretic studies are still lacking and the authors' future work should contribute in that area.

REFERENCES

[1] D. Lirman, N. R. Gracias, B. E. Gintert, A. C. R. Gleason, R. P. Reid, S. Negahdaripour, and P. Kramer, "Development and application of a video-mosaic survey technology to document the status of coral reef communities," *Environmental Monitoring and Assessment*, vol. 125, no. 1-3, pp. 59–73, 2007.

[2] F. Ferreira, G. Veruggio, M. Caccia, E. Zereik, and G. Bruzzone, "A real-time mosaicking algorithm using binary features for ROVs," in *21st Mediterranean Conference on Control and Automation*. IEEE, June 2013, pp. 1267–1273. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6608882

[3] F. Ferreira, G. Veruggio, M. Caccia, and G. Bruzzone, "A survey on real-time motion estimation techniques for underwater robots," *Journal of Real-Time Image Processing*, vol. 11, no. 4, pp. 693–711, 2016. [Online]. Available: http://dx.doi.org/10.1007/s11554-014-0416-z

[4] H. Johannsson, M. Kaess, B. Englot, F. Hover, and J. Leonard, "Imaging sonar-aided navigation for autonomous underwater harbor surveillance," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, Oct. 2010, pp. 4396–4403. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5650831

[5] B. P. Foley, K. Dellaporta, D. Sakellariou, B. S. Bingham, R. Camilli, R. M. Eustice, D. Evagelistis, V. L. Ferrini, K. Katsaros, D. Kourkoumelis, A. Mallios, P. Micha, D. A. Mindell, C. Roman, H. Singh, D. S. Switzer, and T. Theodoulou, "The 2005 Chios ancient shipwreck survey: new methods for underwater archaeology," *Hesperia*, vol. 78, no. 2, pp. 269–305, 2009. [Online]. Available: http://www.jstor.org/pss/25622694

[6] B. Ouyang, F. Dalgleish, S. Negahdaripour, and A. Vuorenkoski, "Experimental study of underwater stereo via pattern projection," in *OCEANS 2012 MTS/IEEE: Harnessing the Power of the Ocean*, 2012.

[7] F. Ferreira, V. Djapic, M. Micheli, and M. Caccia, "Forward looking sonar mosaicing for mine countermeasures," *Annual Reviews in Control*, vol. 40, pp. 212 – 226, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S136757881500053X

[8] M. D. Aykin and S. Negahdaripour, "On feature extraction and region matching for forward scan sonar imaging," in *2012 Oceans*. IEEE, Oct. 2012, pp. 1–9. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6404983

[9] N. Hurtos, N. Palomeras, A. Carrera, M. Carreras, C. P. Bechlioulis, G. C. Karras, S. Hesmati-alamdari, and K. Kyriakopoulos, "Sonar-based chain following using an autonomous underwater vehicle," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, Sept. 2014, pp. 1978–1983. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6942825

[10] http://www.codaoctopus.com/products/echoscope.

[11] http://www.tritech.co.uk/product/multibeam-sonar-for-3d-model-view-of-sonar-imagery-eclipse.

[12] http://http://www.blueview.com/products/3d-multibeam-scanning-sonar/3/.

[13] G. J. Dobeck, "Algorithm fusion for the detection and classification of sea mines in the very shallow water region using side-scan sonar imagery," vol. 4038, 2000, pp. 348–361. [Online]. Available: http://dx.doi.org/10.1117/12.396262

[14] A. Pleskachevsky, S. Lehner, T. Heege, and C. Mott, "Synergy and fusion of optical and synthetic aperture radar satellite data for underwater topography estimation in coastal areas," *Ocean Dynamics*, vol. 61, no. 12, pp. 2099–2120, 2011. [Online]. Available: http://dx.doi.org/10.1007/s10236-011-0460-1

[15] D. Loebis, R. Sutton, and J. Chudley, “Review of multisensor data fusion techniques and their application to autonomous underwater vehicle navigation,” *Journal of Marine Engineering & Technology*, vol. 1, no. 1, pp. 3–14, 2002. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/20464177.2002.11020159

[16] M. Andersson, S. Ntalampiras, T. Ganchev, J. Rydell, J. Ahlberg, and N. Fakotakis, “Fusion of acoustic and optical sensor data for automatic fight detection in urban environments,” in *Information Fusion (FUSION), 2010 13th Conference on*, July 2010, pp. 1–8.

[17] L. Grajales and I. V. Nicolaescu, “Wearable multisensor heart rate monitor,” *Wearable and Implantable Body Sensor Networks, International Workshop on*, vol. 0, pp. 154–157, 2006.

[18] M. Penza, G. Cassano, P. Aversa, A. Cusano, A. Cutolo, M. Giordano, and L. Nicolais, “Carbon nanotube acoustic and optical sensors for volatile organic compound detection,” *Nanotechnology*, vol. 16, no. 11, p. 2536, 2005. [Online]. Available: http://stacks.iop.org/0957-4484/16/i=11/a=013

[19] T. Smith, “Weapon location by acoustic-optic sensor fusion,” Sept. 16 2003, uS Patent 6,621,764. [Online]. Available: https://www.google.com/patents/US6621764

[20] K. Bates and G. Vardi, “Systems and methods for minimally-invasive optical-acoustic imaging,” July 17 2007, uS Patent 7,245,789. [Online]. Available: https://www.google.com/patents/US7245789

[21] M. J. Chantler, D. B. Lindsay, C. S. Reid, and V. J. C. Wright, “Optical and acoustic range sensing for underwater robotics,” in *OCEANS '94. 'Oceans Engineering for Today's Technology and Tomorrow's Preservation.' Proceedings*, vol. 1, Sep, pp. I/205–I/210 vol.1.

[22] S. Bejarano, P. J. Mumby, J. D. Hedley, and I. Sotheran, “Combining optical and acoustic data to enhance the detection of caribbean forereef habitats,” *Remote Sensing of Environment*, vol. 114, no. 11, pp. 2768 – 2778, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0034425710002026

[23] E. Fumagalli, M. Bibuli, M. Caccia, E. Zereik, F. D. Bianco, L. Gasperini, G. Stanghellini, and G. Bruzzone, “Combined acoustic and video characterization of coastal environment by means of unmanned surface vehicles,” {*IFAC*} *Proceedings Volumes*, vol. 47, no. 3, pp. 4240 – 4245, 2014, 19th {IFAC} World Congress. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1474667016422676

[24] A. Lagudi, G. Bianco, M. Muzzupappa, and F. Bruno, “An alignment method for the integration of underwater 3d data captured by a stereovision system and an acoustic camera,” *Sensors*, vol. 16, no. 4, p. 536, 2016. [Online]. Available: http://www.mdpi.com/1424-8220/16/4/536

[25] J. Potter, A. Pack, M. Hoffman-Kuhnt, T. B. Koay, P. Seekings, and M. Chitre, “A synchronised acoustic array, rangefinder & video system with examples from 'singing' humpback whales (megaptera noveangliae),” in *Proceedings of the European Cetacean Society Conference*, Canary Islands, April 2007.

[26] J. Potter, A. Pack, J. Reidenberg, M. Hoffman-Kuhnt, P. Seekings, M. Chitre, T. B. Koay, and L. Herman, “Humpback whale song source location in the head, source levels and directionally from in-situ rebreather diver recordings,” in *17th Biennal conference on the biology of marine mammals*, P. Best and e. M. Bester, Eds., Cape Town, 2007.

[27] S. Majumder, S. Scheding, and H. F. Durrant-Whyte, “Multisensor data fusion for underwater navigation,” *Robotics and Autonomous Systems*, vol. 35, no. 2, pp. 97–108, 2001.

[28] S. Williams and I. Mahon, “Simultaneous localisation and mapping on the great barrier reef,” in *Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on*, vol. 2, April 2004, pp. 1771–1776 Vol.2.

[29] H. Singh, C. Roman, L. Whitcomb, and D. Yoerger, “Advances in Fusion of High Resolution Underwater Optical and Acoustic Data,” in *2000 International Symposium on Underwater Technologies UT 00*, 2000, pp. 206–211.

[30] G. Sulzberger, J. Bono, R. J. Manley, T. Clem, L. Vaizer, and R. Holtzapple, “Hunting sea mines with uuv-based magnetic and electro-optic sensors,” in *OCEANS 2009*, Oct 2009, pp. 1–5.

[31] J. S. N. Hurtós, X. Cufí, “Integration of optical and acoustic sensors for 3D underwater scene reconstruction,” 2009. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.298.8035

[32] C. G. Kunz, “Autonomous underwater vehicle navigation and mapping in dynamic, unstructured environments,” PhD, MIT and WHOI, 2012.

[33] G. Inglis, “Hybrid Optical Acoustic Seafloor Mapping,” Ph.D. dissertation, 2013. [Online]. Available: http://digitalcommons.uri.edu/oa_diss/64

[34] D. Moroni, M. A. Pascali, M. Reggiannini, and O. Salvetti, “Underwater scene understanding by optical and acoustic data integration,” *Proceedings of Meetings on Acoustics*, vol. 17, no. 1, 2013. [Online]. Available: http://scitation.aip.org/content/asa/journal/poma/17/1/10.1121/1.4792225

[35] B. Kalyan and A. Balasuriya, “Multisensor data fusion approach for terrain aided navigation of autonomous underwater vehicles,” in *Oceans '04 MTS/IEEE Techno-Ocean '04 (IEEE Cat. No.04CH37600)*, vol. 4. IEEE, 2004, pp. 2013–2018. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1406452

[36] F. S. Hover, R. M. Eustice, A. Kim, B. Englot, H. Johannsson, M. Kaess, and J. J. Leonard, “Advanced Perception , Navigation and Planning for Autonomous In-Water Ship Hull Inspection,” *International Journal of Robotic Research*, vol. 31, no. 12, pp. 1445–1464, 2012.

[37] D. Krout, G. Okopal, and E. Hanusa, “Video data and sonar data: Real world data fusion example,” in *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*. Chicago: IEEE, 2011, pp. 1–5. [Online]. Available: http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=5977701

[38] S. Negahdaripour, “Epipolar geometry of opti-acoustic stereo imaging.” *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 10, pp. 1776–88, Oct. 2007. [Online]. Available: http://www.researchgate.net/publication/6139650_Epipolar_geometry_of_opti-acoustic_stereo_imaging

[39] S. Negahdaripour, H. Sekkati, and H. Pirsiavash, “Opti-acoustic stereo imaging: On system calibration and 3-D target reconstruction,” *IEEE Transactions on Image Processing*, vol. 18, no. 6, pp. 1203–1214, 2009.

[40] S. Negahdaripour, “A new method for calibration of an opti-acoustic stereo imaging system,” in *OCEANS 2010 MTS/IEEE SEATTLE*. IEEE, Sept. 2010, pp. 1–7. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5664563

[41] S. Negahdaripour and A. Taatian, “3-D motion and structure estimation for arbitrary scenes from 2-D optical and sonar video,” in *OCEANS 2008*. IEEE, 2008, pp. 1–8. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5151985

[42] M. Babaee and S. Negahdaripour, “3-D object modeling from 2-D occluding contour correspondences by opti-acoustic stereo imaging,” *Computer Vision and Image Understanding*, vol. 132, pp. 56–74, Mar. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1077314214002112

[43] ——, “3-D object modeling from 2-D occluding contour correspondences by opti-acoustic stereo imaging,” *Computer Vision and Image Understanding*, vol. 132, pp. 56–74, 2015. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1077314214002112

[44] A. Fusiello and V. Murino, “Augmented scene modeling and visualization by optical and acoustic sensor integration.” *IEEE transactions on visualization and computer graphics*, vol. 10, no. 6, pp. 625–36, Jan. 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15527045

[45] P. Drap, D. Merad, J.-M. Boï, W. Boubguira, A. Mahiddine, B. Chemisky, E. Seguin, F. Alcala, and O. Bianchimani, “Rov-3d: 3d underwater survey combining optical and acoustic sensor,” in *Proceedings of the 12th International Conference on Virtual Reality, Archaeology and Cultural Heritage*, ser. VAST'11. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2011, pp. 177–184. [Online]. Available: http://dx.doi.org/10.2312/VAST/VAST11/177-184

[46] A. Mahiddine, R. Iguernaissi, D. Merad, P. Drap, and y. Jean-Marc Boï, booktitle=ICPRAM, “3d registration of multi-modal data using surface fitting.”

[47] S. Negahdaripour, “Calibration of DIDSON Forward-Scan Acoustic Video Camera,” in *Proceedings of OCEANS 2005 MTS/IEEE*. IEEE, 2005, pp. 1–8. [Online]. Available: http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=1639932

Document Data Sheet

Security Classification		*Project No.*
Document Serial No. CMRE-PR-2019-078	*Date of Issue* June 2019	*Total Pages* 6 pp.

Author(s)

Fausto Ferreira, Diogo Machado, Gabriele Ferri, Samantha Dugelay, John Potter

Title

Underwater optical and acoustic imaging: A time for fusion? a brief overview of the state-of-the-art

Abstract

Underwater optical imaging has several drawbacks inherent to the physical medium such as light attenuation and turbidity. Sonars try to obviate those issues although, typically, they have lower resolutions. Combining visual and sonar data in underwater applications is still not popular, but researchers' interest in the topic is growing. However, with the advent of recent higher resolution sonar systems, the approach of combining/fusing information from both sensory modalities can bring improvements to underwater imaging. This has special interest for applications such as autonomous navigation, mapping and object recognition. In this paper we investigate the state of the art for these systems and present the most relevant approaches found in the literature.

Keywords

Optical sensors, optical imaging, sonar, acoustics, cameras, adaptive optics, optical attenuators

Issuing Organization

NATO Science and Technology Organization
Centre for Maritime Research and Experimentation

Viale San Bartolomeo 400, 19126 La Spezia, Italy

[From N. America:
STO CMRE
Unit 31318, Box 19, APO AE 09613-1318]

Tel: +39 0187 527 361
Fax:+39 0187 527 700

E-mail: library@cmre.nato.int