ENGN4627 Research Report: Simultaneous Enhancement and Super-Resolution of Underwater Imagery for Improved Visual Perception

Executive Summary

The paper presented to the Robotics Science and Systems Conference is on the simultaneous enhancement and super resolution (SESR) of underwater imagery [1]. The problems surrounding underwater imagery are introduced in the report, identifying the need for underwater-specific presented contextualizing the field in which the research paper was presented. Md Jahidul Islam, Peigen Luo and Junaed Sattar's paper presents the Deep SESR algorithm which uses a machine learning approach to provide SESR with near-real-time capabilities. The SESR approach is able to outperform the existing state-of-the-art (SOTA) systems in both image enhancement and super-resolution in underwater environments marking a clear progression in this field. The system's robustness also makes it competitive results on terrestrial images and strong generalized performance on other natural images. The Deep SESR approach also includes saliency prediction to be used by underwater vehicles and the system's compact architecture allows for application by real-time dependent systems with relatively low computational requirements. The paper also notably includes the UFO-120 database, which was developed and published to train future underwater SESR algorithms. The applications for this technology are explored, with notable developments made aiding autonomous underwater vehicles (AUV) in either improved performance or as a cost-effective solution to more expensive equipment. The implications of each support both social causes such as environmental monitoring or commercial interests in mapping and infrastructure inspection. The findings presented by the paper are significant in the improvement of underwater computer vision. Deep SESR not only represents improvements in image quality techniques but the ability to meaningfully apply them to visually dependent robotic systems for significant social and commercial applications.

Background

Computer vision applications in underwater environments face additional challenges compared to applications in other environments. The environmental context leads to a variety of optical distortions and noise that make underwater visualization a challenging task requiring revised image formation models and extensive corrections. Lighting is a particular issue in underwater environments. Light can be severely attenuated by the water and small angle scattering [2] can produce beams of light in the water that are non-linear in nature. This inconsistent lighting frequently leads to low contrast and image blur. Water also absorbs different frequencies of light at different rates and so images also face hue distortions as particular frequencies are absorbed by the water. Systems providing their own lighting risk introducing backscatter, a phenomenon where particles in the water light up [3] as well as the target feature. This frequently is modelled as noise but can also distort brightness characteristics of the image. Models of underwater cameras can also face refractive challenges with lenses and camera design. As demonstrated by Figure 1, these challenges can have significant impact on the image characteristics and present the need for image enhancement for improved perceptual qualities of inputs to systems.

Figure 1: Comparison of image with optical distortions to high resolution ground truth

The field of underwater image enhancement is well researched with many competing systems developed to address these problems. These solutions are generally categorized as physics-based models and learning-based models. Physics-based models aim to estimate the degradation features and then remove them through corrections of this estimation [1]. The drawback of these approaches is that they are computationally expensive and frequently rely on additional information, such as scene depth, that is not always available in applications. There have been gradual improvements in this field however more recent study has shifted towards learning-based approaches. Unsupervised color correction method (UCM) developed in 2010 at Coventry University [4] remains among the most effective. Multi-scale models such as MS-Fusion developed in 2012 at Hasselt University [5] have made minor improvements while the leading physics-based model appears to be MS-Retinex. MS- Retinex is a modified multi-scale approach developed in 2017 between researchers from the University of Portsmouth and Ocean University of China [6]. Learning-based approaches use convolutional neural networks (CNNs) and generative adversarial networks (GANs) to approximate the problem's solution relying purely on RGB image data. These learning-based solutions are found to generally have superior performance on test sets [1] without additional measures and generally are less computationally expensive once established. Research in this area is relatively more recent and developing rapidly. Md Jahidul Islam's research at the University of Minnesota appears to lead in the development of this area. The FUnIE-GAN developed in 2019 is among the best in underwater enhancement while UGAN developed in 2018 is also competitive. Another notable algorithm, Fusion-GAN, was published in 2019 from Nanjing University of Information Science and Technology and has similar performance to UGAN [7].

The automatic generation of high-resolution images from low resolution measurements is also a well-developed problem in robotics and computer vision. Visually-guided robots frequently use the HR capability to 'zoom-in' on particular regions of interest (RoIs) for detailed perception [1]. This increased perception of RoIs can inform the system to make navigational or other operational decisions based on the region. This field has been well studied with many competing algorithms for both terrestrial and underwater applications. Currently the SOTA terrestrial models are considered to be SRCNN developed in 2015 by the University of Hong Kong [8] and the SRGAN model developed in 2017 at the Imperial College London [9]. Islam's research at the University of Minnesota again appears to be leading the field in algorithms specifically designed for underwater applications. The university's SRDRM and SRDRM-GAN models are among the leading algorithms along with RSRGAN [10] developed recently at the Ocean University of China.

Saliency refers to identifying what is important or noticeable in an image. From a computer vision perspective this is the ability to find relevant regions in the image. Stimulus-driven approaches to saliency prediction use image features to quantify feature contrast within a scene [1]. An alternative approach is the goal-driven approach where the focus is primarily on finding image regions that best match specific queries [1]. In the context of the research paper, saliency prediction primarily focuses on identifying relevant foreground regions using a stimulus-driven approach. While not directly supporting development in image enhancement, saliency can facilitate faster search methods and scene understanding. Saliency prediction can be used to identify RoIs such that SESR need only be applied to particular areas rather than the whole image.

The findings of Islam's prior research into both image enhancement and super-resolution through machine learning techniques clearly influenced his approach to SESR. The improved quality and lower computational costs of machine learning approaches in both image enhancement and SISR compared to existing techniques encouraged its further development. Although approaches to these individual problems were successful, the applications of each identified the need for a unified system capable of achieving both approaches efficiently. The paper also draws on previous findings such as training instability in GANs, successes in residual-in-residual learning and research in machine learning cost functions to better develop machine learning approaches.

There are a number key underlying assumptions made by the paper. Firstly, the paper assumes that the techniques used to generate the UFO-120 dataset accurately model real-world distortions created by the environment. An existing network (CycleGAN [11]) was used to mimic distortions. Limitations in this approach could easily be transferred to the SESR algorithm where it might learn to correct CycleGAN's approach to noise generation rather than realistic modelling. Secondly, the paper assumes the original image dataset is without distortion. Although the original dataset images are HR, significant distortions in this dataset would limit the model's learning capabilities as it optimized to include these. This concern is not directly addressed and it is assumed that original images are without any distortion.

The main research groups in the field of underwater computer vision appear to be based in the University of Minnesota and Ocean University of China. The University of Minnesota is an unusual area for this research to be conducted as the state is landlocked. However, the University's Interactive Robotics and Vision Laboratory run by Junaed Sattar is centered around solving a range of underwater based robotics problems. The department's specialization in underwater vision makes it a lead researcher in this field. There are no clear links to funding or support from organizations outside of the University of Minnesota. The Ocean University of China has strong connections to the Chinese government and support a wide variety of research in many ocean-based fields. The university specialized primarily on marine sciences and marine-based engineering. Their goal is to be "implementing a strategy to make China strong by further developing the ocean." [12] Improved underwater computer vision capabilities is one aspect of a broad range of University research.

Contribution

The paper makes a number of key contributions that are important to the field of study, both conceptually and in applications. The paper presents a learning-based solution for SESR that is efficient and able to be applied on a single-board computer with low memory requirements. The solution performs image enhancement and super-resolution at levels competing with, or exceeding, other SOTA solutions in those fields. The solution can also be generalized to provide competitive performance in terrestrial settings and other environments.

Simultaneous Enhancement and Super-Resolution

The paper presents the first unified approach for SESR with an end-to-end trainable model and at the time of publishing, is the only fully integrated SESR method. The problem is defined as learning a pixel-to-pixel mapping from inputs to higher-resolution output as a generative function. This was extended by including saliency prediction over the feature space. The problem is posed as the learning of the generative function:

$$G: X \rightarrow S, E, Y \quad [1]$$

where S, E and Y refer to the predicted saliency map, enhanced image (at resolution of input) and high-resolution enhanced image. This was achieved by developing a Deep SESR model using residual dense blocks (RDBs), a feature extraction network (FENet) and an auxiliary attention network (AAN) under an end-to-end architecture. The architecture is shown in Figure 2 and component used are defined in the paper [1].

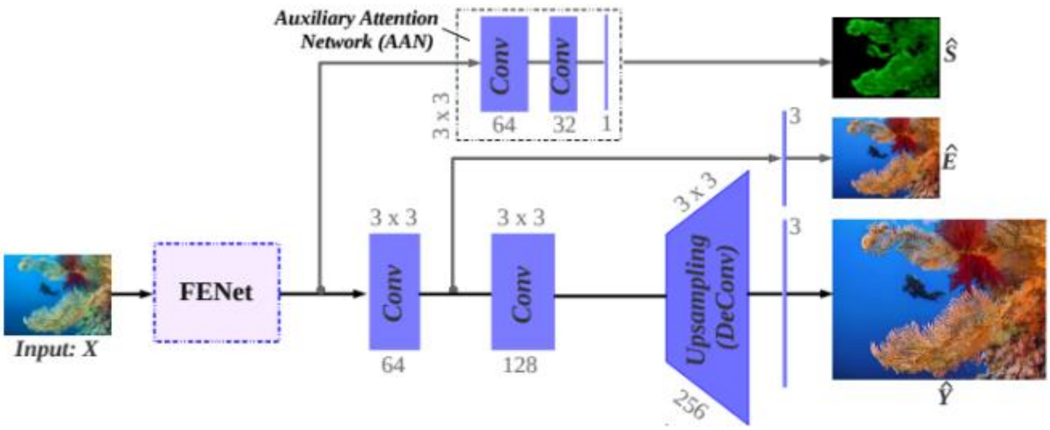

Figure 2: Deep SESR complete system architecture [1]

The unified approach conceptually makes it well suited to real world applications. The unified architecture reduces memory requirements and running different aspects of the program concurrently allows for faster run-time. In contrast, to achieve similar results with alternatives in enhancement and super-resolution (but not saliency), 2 separate programs would be required, needing more memory and having lower run-times. Deep SESR requires only 10MB of memory and can run at 7.75 FPS on a single-board computer [1].

This unified architecture approach allows for the Deep SESR to address multiple important aspects of the outputs from the generative function in the loss function. The authors gave a detailed description of how the loss function is derived and evaluated. In particular, the authors presented new approaches to exploit saliency prediction for improved loss functions and refined existing loss functions in literature to ones that produced better results. The unified approach allows Deep SESR to address multiple aspects of the visualization problem to balance outcomes. More limited approaches may be able to achieve improved performance in specific aspects but this may be at the loss of other aspects, such as saliency prediction, leading to greater error when applied. The well-defined loss functions presented also allow those applying Deep SESR to modify aspects for improved performance in areas important to that application. The Deep SESR current applies a basic minimum value function of the weighted sum of all loss functions. For example, an application such as coral reef monitoring with greater need for color information could manipulate this final loss function to improve performance in a particular area. The unified SESR approach is particularly important as it is also able to outperform competing SOTA systems in both image enhancement and single image super-resolution.

Table 1: Quantitative performance comparison for enhancement (Best Scores -Red, 2nd Best Scores-Blue) [1]

	Dataset	RGHS	UCM	MS-Fusion	MS-Retinex	Water-Net	UGAN	Fusion-GAN	FUnIE-GAN	**Deep SESR**
PSNR	UFO-120	20.05 ± 3.1	20.99 ± 2.2	21.32 ± 3.3	21.69 ± 3.6	22.46 ± 1.9	23.45 ± 3.1	24.07 ± 2.1	25.15 ± 2.3	27.15 ± 3.2
	EUVP	20.12 ± 2.9	20.55 ± 1.8	19.85 ± 2.4	21.27 ± 3.1	20.14 ± 2.3	23.67 ± 1.5	23.77 ± 2.4	26.78 ± 1.1	25.25 ± 2.1
	UImNet	19.98 ± 1.8	20.48 ± 2.2	19.59 ± 3.2	22.63 ± 2.5	21.02 ± 1.6	23.88 ± 2.1	23.12 ± 1.9	24.68 ± 2.4	25.52 ± 2.7
SSIM	UFO-120	0.75 ± 0.06	0.78 ± 0.07	0.79 ± 0.09	0.75 ± 0.10	0.79 ± 0.05	0.80 ± 0.08	0.82 ± 0.07	0.82 ± 0.08	0.84 ± 0.03
	EUVP	0.69 ± 0.11	0.73 ± 0.14	0.70 ± 0.05	0.69 ± 0.15	0.68 ± 0.18	0.67 ± 0.11	0.68 ± 0.05	0.86 ± 0.05	0.75 ± 0.07
	UImNet	0.61 ± 0.08	0.67 ± 0.06	0.64 ± 0.11	0.74 ± 0.04	0.71 ± 0.07	0.79 ± 0.08	0.75 ± 0.07	0.77 ± 0.06	0.81 ± 0.05
UIQM	UFO-120	2.36 ± 0.33	2.41 ± 0.53	2.76 ± 0.45	2.69 ± 0.59	2.83 ± 0.48	3.04 ± 0.28	2.98 ± 0.28	3.09 ± 0.51	3.13 ± 0.45
	EUVP	2.45 ± 0.46	2.48 ± 0.77	2.51 ± 0.36	2.48 ± 0.09	2.55 ± 0.06	2.70 ± 0.31	2.58 ± 0.07	2.95 ± 0.38	2.98 ± 0.28
	UImNet	2.32 ± 0.48	2.38 ± 0.42	2.79 ± 0.55	2.84 ± 0.37	2.92 ± 0.35	3.32 ± 0.55	3.19 ± 0.27	3.23 ± 0.32	3.26 ± 0.36

Table 1 shows that Deep SESR ranks as best or second best in image enhancement for all training sets when compared to other SOTA enhancement methods. Table 2, also shows that Deep SESR ranks as best for all bar 2 cases in super-resolution

Table 2: Quantitative performance comparison for Super-Resolution (Best Scores-Red, 2nd Best Scores-Blue) [1]

		PSNR			*SSIM*			*UIQM*		
	Model	2×	3×	4×	2×	3×	4×	2×	3×	4×
UFO-120	SRCNN	24.75 ± 3.7	22.22 ± 3.9	19.05 ± 2.3	$.72 \pm .07$	$.65 \pm .09$	$.56 \pm .12$	2.39 ± 0.35	2.24 ± 0.17	2.02 ± 0.47
	SRResNet	25.23 ± 4.1	23.85 ± 2.8	19.13 ± 2.4	$.74 \pm .08$	$.68 \pm .07$	$.56 \pm .05$	2.42 ± 0.37	2.18 ± 0.26	2.09 ± 0.30
	SRGAN	26.11 ± 3.9	23.87 ± 4.2	21.08 ± 2.3	$.75 \pm .06$	$.70 \pm .05$	$.58 \pm .09$	2.44 ± 0.28	2.39 ± 0.25	2.26 ± 0.17
	RSRGAN	25.25 ± 4.3	23.15 ± 4.1	20.25 ± 2.4	$.79 \pm .08$	$.71 \pm .08$	$.58 \pm .04$	2.41 ± 0.29	2.38 ± 0.31	2.27 ± 0.22
	SRDRM	26.23 ± 4.4	⊘	22.26 ± 2.5	$.79 \pm .09$	⊘	$.59 \pm .05$	2.45 ± 0.43	⊘	2.28 ± 0.35
	SRDRM-GAN	26.26 ± 4.3	⊘	22.21 ± 2.4	$.78 \pm .08$	⊘	$.58 \pm .13$	2.42 ± 0.30	⊘	2.27 ± 0.44
	Deep SESR	28.57 ± 3.5	26.86 ± 4.1	24.75 ± 2.8	$.85 \pm .09$	$.75 \pm .06$	$.66 \pm .05$	3.09 ± 0.41	2.87 ± 0.39	2.55 ± 0.35
USR-248	SRCNN	24.88 ± 4.4	24.01 ± 3.5	23.75 ± 3.2	$.73 \pm .08$	$.70 \pm .10$	$.69 \pm .12$	2.38 ± 0.38	2.31 ± 0.29	2.21 ± 0.68
	SRResNet	24.96 ± 3.7	23.39 ± 5.2	22.21 ± 3.6	$.74 \pm .07$	$.71 \pm .11$	$.70 \pm .08$	2.42 ± 0.48	2.33 ± 0.58	2.27 ± 0.70
	SRGAN	25.76 ± 3.5	25.02 ± 3.9	24.36 ± 4.3	$.77 \pm .06$	$.75 \pm .05$	$.69 \pm .13$	2.53 ± 0.42	2.65 ± 0.44	2.75 ± 0.66
	RSRGAN	25.11 ± 2.9	24.96 ± 4.7	24.15 ± 2.9	$.75 \pm .06$	$.72 \pm .09$	$.71 \pm .09$	2.42 ± 0.35	2.49 ± 0.56	2.55 ± 0.47
	SRDRM	26.16 ± 3.5	⊘	24.96 ± 3.3	$.77 \pm .10$	⊘	$.72 \pm .11$	2.47 ± 0.69	⊘	2.35 ± 0.51
	SRDRM-GAN	26.77 ± 4.1	⊘	24.77 ± 3.4	$.82 \pm .07$	⊘	$.70 \pm .12$	2.87 ± 0.55	⊘	2.81 ± 0.56
	Deep SESR	27.03 ± 2.9	25.92 ± 3.5	24.59 ± 3.8	$.88 \pm .05$	$.76 \pm .05$	$.71 \pm .08$	3.15 ± 0.44	3.04 ± 0.37	2.96 ± 0.28

The results in both table 1 and table 2 reflect that the Deep SESR is capable of outperforming SOTA systems marking a clear progression in the field of underwater visualization. Notably, it is able to outperform competing algorithms on a variety of training sets and multiple performance metrics reflecting a robust system performance.

UFO-120 Dataset

The paper makes a significant contribution to the field of underwater visualization by developing and providing the UFO-120 dataset. This is the first dataset to facilitate large-scale SESR learning, providing opportunities for other researchers to train and test competing algorithms to further develop the field.

The dataset contains 1500 annotated samples with a test set of 120 samples. The dataset was developed by taking high resolution images from ocean explorations in a variety of locations and water types. The low-resolution images were developed from these high-resolution images through standard optical/spatial image degradation [1] technics. These techniques utilized a GAN (GAN) model to generate distorting that mimicked the characteristics of an underwater environment and then down-sampled using gaussian blurring [11]. This approach easily allows machine learning algorithms to interpret the inputs and compare performances of outputs to the original images. The saliency maps were manually annotated. This dataset is the first to publicly include ground truth saliency maps meaning that it is the only dataset were machine learning can be applied to learn saliency in this environmental setting. While the dataset was generated using existing methods and technics, its generation enables others to train and test their own approaches to similar problems. It makes particularly valuable contributions by including saliency maps so that this aspect can be learnt alongside image enhancement and super-resolution of improved overall performance.

Social/commercial Impact

Autonomous Underwater Vehicles (AUVs) and Remote Operated Underwater Vehicles (ROVs) are widely used in a number of applications. Diving is an operation that always contains high risks, especially in the ocean where conditions may change rapidly. AUVs and ROVs can be applied to reduce risk to human life and in some circumstances can even out-perform human-counterparts. The near real-time capabilities of Deep SESR mean that it can be applied to these cases where non real-time enhancement or SISR systems would not be appropriate. The findings and presented solution discussed in the paper could have wide reaching impacts in underwater visualization applications as with AUVs and ROVs. The research and technology presented in this paper suggest that Deep SESR is suitable for near real-time use by visually guided underwater robots. Deep SESR's high performance in both enhancement and super-resolution provides improved quality of inputs for underwater systems. This also has applications in areas where high-resolution is not a strong priority. The SESR system can reduce the camera quality and resolution requirements of systems to function effectively. Systems that had minimum resolution requirements can transition to lower resolution cameras with Deep SESR applied. This could result in substantial costs being saved.

Saliency prediction could be especially helpful to these automated systems understanding the positioning of the environment. The ability to recognize that an object is relatively close can help algorithms better avoid collisions and choose optimal paths. This capability already exists [13] however there is not an exiting solution that combines this with image enhancement or super resolution.

An application of AUVs are in monitoring marine species and coral reef areas. The vehicles create relatively little disturbance in the environment and can potentially cover far greater area than divers [14]. Deep SESR can improve performances of AUVs in this application by reducing camera costs, and improve path finding and localization through improved inputs. The improved image quality can also be used to more accurately identify marine species and coral health. This allows for workers monitoring marine species and reefs to have improved information and greater understanding of conservation and the environment.

AUV and ROVs are also active in industrial settings. AUVs have been used to make detailed maps of the sea floor with applications for scientific research and oil and gas drilling [15]. These vehicles are also used to perform inspections on man-made underwater infrastructure [15]. Application of Deep SESR could be used to improve the visual odometry and localization of the mapping vehicles resulting in more accurate mapping of the seafloor environment. This better informs design decisions and allows for greater capability in planning. Improved image qualities will also make it easier for engineers to identify flaws in infrastructure upon inspection. This reduces risks associated with failure of such system as failure or weaknesses can be identified early where there are lower consequences. Given the enormous environment and economic cost of failure in oil or gas industries, even marginal improvements in risk assessments can have significant commercial impact.

Deep SESR's reasonable performance on terrestrial images demonstrate that the proposed architecture has the capacity to learn for more generalized applications. Therefore, with further development this approach could be applied to other autonomous technology such as UAVs or drones to improve visual inputs. Although there is lower demand in this area due to less problematic distortions, the technology presents value through costs saved due to lower camera resolution requirements. The strong terrestrial performance by the UFO-120 trained Deep SESR indicates potential applications in technology transitioning between dry and underwater environments.

Specifically, within the University of Minnesota this research can further the research of others in the area. This improved underwater visualization could be applied to research on improving the detection of marine debris with the goal to remove pollution from the ocean, research in improving the semantic segmentation, underwater human-robot collaboration (such as human-robot communication through underwater hand signals) and research in AUV localization techniques. Supporting research in each of these fields could have further social and commercial impacts as AUV and ROV tooling develops [16].

Bibliography

[1] J. Islam, P. Luo and J. Sattar, "Robotics: Science and Systems," 12 July 2020. [Online]. Available: https://roboticsconference.org/program/papers/18/. [Accessed 10 2020].

[2] L. Mullen, "spiedigitallibrary.org," 2 May 2009. [Online]. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/7324/1/Optical-propagation-in-the-underwater-environment/10.1117/12.820205.full?SSO=1&tab=ArticleLink. [Accessed 10 2020].

[3] S. Gietler, "UW Photography," 2020. [Online]. Available: https://www.uwphotographyguide.com/backscatter-underwater. [Accessed 10 2020].

[4] K. Iqbal, M. Odetayo, A. James, R. A. Sakam and A. Z. H. Talib, "Enhancing the Low Quality Images Using Unsupervised Colour Correction Method," *IEEE Inernational Conference on Systems, Man and Cybernetics,* pp. 1703-1709, 2010.

[5] C. Ancuti, C. O. Ancutti, T. Haber and P. Bekaert, "Enhancing Underwater Images and Videos by Fusion," *IEEE Conference on Computer Vision and Pattern Recognition,* pp. 81-88, 2012.

[6] S. Zhang, T. Wang, J. Dong and H. Yu, "Underwater Image Enhancement via extended multi-scale retinex," *Neurocomputing,* pp. 245:1-9, 2017.

[7] H. Li, J. Li and W. Wang, "A fusion adversarial underwater image enhancement network with a public test dataset," arXiv, 2019.

[8] C. Dong, C. C. Loy, K. He and X. Tang, "Image Super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and machine intelligence,* vol. 38, no. 2, pp. 295-307, 2015.

[9] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta and A. Aitken, "Photo-Realistic Single Image Super-resolution using a generative adversarial network," IEEE, 2017.

[10] Y. Chen, J. Sun, W. Jiao and G. Zhong, "Recovering Super-resolution generative adversarial network for underwater images," in *Neural Information Processing*, 2019.

[11] J.-Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *IEEE International Conference on Computer Vision*, 2017.

[12] Ocean University of China, “Ocean University of China,” 2 11 2020. [Online]. Available: http://eweb.ouc.edu.cn/visionwmission/list.htm.

[13] L. Zhang, B. He, Y. Song and T. Yan, “Underwater image feature extraction and matching based on visual saliency detection,” *IEEEL OCEANS,* pp. 1-4, 2016.

[14] “Automatic annotation of coral reefs using deep learning,” *IEEE OCEANs,* 2016.

[15] D. Bingham, T. Drake, A. Hill and R. Lott, “The application of autonomous underwater vehicle (AUV) technology in the oil industry- Vision and experiences,” FIG International Congress, Washington DC., 2002.

[16] “Interactive Robotics and Vision Lab,” University of Minnesota, 1 11 2020. [Online]. Available: http://irvlab.cs.umn.edu/.