Underwater Object Tracking Benchmark and Dataset

Landry Kezebou, Student Member, IEEE,
Victor Oludare, Student Member, IEEE,
Karen Panetta, Fellow, IEEE
Department of Electrical and Computer Engineering
Medford, MA, USA
Landry.Kezebou@tufts.edu, Victor.Oludare@tufts.edu,
Karen@ece.tufts.edu

Sos S Agaian, Fellow, IEEE,
Department of Computer Science
The City University of New York,
New York City, NY, USA
sos.agaian@csi.cuny.edu

***Abstract*—While there has been tremendous advancement in object tracking for open air visual data, much less work has been done for underwater object tracking. This is due to the low quality of underwater visual data. Underwater visual data suffers distortions in contrast and sharpness, as a result of refraction and absorption of light, and particles, which all vary dependent on the depth, color and nature of water. Although there currently exists several object tracking algorithms with proven record of high speed, precision and success rate, these algorithms work best for open air tracking, and considerably degrade in performance when tracking targets in underwater environments, as it is presented in this paper. The advancement made in open air tracking has been facilitated by availability of multiple benchmark and dataset. However, no such benchmark and dataset exist for underwater tracking, and this lack of data has hindered development of dedicated underwater tracking algorithms. In this paper, we present: a) the first underwater tracking benchmark dataset consisting of 32 videos, and a total of 24241 annotated frames, averaging 29.15 seconds and 757.53 frames per video, to help improve underwater tracking; and b) a comparative performance analysis of existing tracking algorithms in underwater environment as opposed to open air.**

***Index Terms*—Underwater visual data, object tracking,benchmark dataset, performance evaluation, image enhancement**

I. Introduction

Object tracking is one of the most important problems in computer vision, and as such has attracted the attention of many researchers in recent years. It finds application in domains such as homeland security, port and marine security, search and rescue operations, disaster recovery, human-computer interaction, video communication and compression, augmented reality, traffic control, medical imaging, and video editing [1]. Here, the focus will be on marine border security applications.

There has been several object tracking benchmarks for both single object and multiple object tracking. The most popular include, the Multiple Object Tracking (MOT) [2], the Visual Object Tracking (VOT) [3], Object Tracking Becnhmark (OTB) [4], and Thermal object Tracking (TOT) benchmark [2]. Competitions such as VOT challenge [3], the MOT challenge [5], NUSPRO VOT challenge [6], and the Thermal Infrared Object Tracking challenge [7], have helped push the boundaries of these benchmarks with consistent improvement in speed, precision and success rate. Unfortunately, all the existing tracking benchmarks datasets focus on open air target tracking. There currently exist no such benchmark or research competition for underwater object tracking. The low quality of underwater visual data, due to distortions such as contrast, absorption, particles, refraction of light, color,and sharpness, has made underwater tracking less attractive to researchers and as such has not received the same attention as open air tracking counterpart. While there has been tremendous advance and success in open air tracking, the existing trackers considerably degrade in performance when tested on underwater visual data, hence the need to create an underwater tracking benchmark.

978-1-7281-5092-5/19/$31.00 ©2019 IEEE

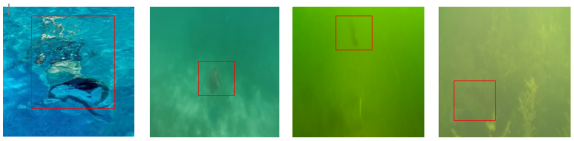

Figure 1. Example of distorted underwater images. Left to right: diver, sea turtle, fish, largemouth bass.

Our main contributions include : a) create a first underwater water object tracking dataset with complete annotation to facilitate bench marking, b) create a first Underwater Object Tracking (UOT) Benchmark by performing a comparative analysis of existing trackers and documenting current performance.

The rest of this paper is structured as follows: section II- background, section III- Dataset , section IV - Evaluation methodology and Results, and section V - Conclusion.

II. Background

The aforementioned research competitions for single and multiple objects tracking, has spurred the development and improvement of several tracking algorithms suitable for variety of applications and most efficient for particular tasks, often trading off speed vs precision and success rate. Tracking algorithms are developed using different methods including correlation filter based tracking [8], target estimation based tracking [9], spatial and intensity information [10], [11].

Some of the popular object tracking algorithms include Kernelized correlation Filter (KCF) [8], TLD tracker [12], Boosting [13], CSRT [14], MIL [15], GOTURN [16], MOSSE [17], MEDIANFLOW [18], and LKT tracker [11]; most of which are all available in OpenCV library. Other state of the art trackers include: ECO [19], CCOT [20], STAPLE [21], STRCF [22], BACF [23], DCF [24], and SAMF [25], whose Matlab based source code are available on Github. The above mentioned trackers are considered state of the art tracking algorithms based on their performance on readily available open air tracking data. We refer the reader to the respective publications and GitHub repositories to learn more about how each of these trackers operate.

There exist a few underwater tracking algorithms using different techniques such as time frequency signatures [26], weighted template matching [27], Kalman filter [28], color based light attenuation [29]. However, these methods do not generalize well and also don't perform comparable to open air state of the art object tracking algorithms. S. Bazeille et al [29] proposed a color based method for detecting and tracking underwater objects using light attenuation and color scheme. The algorithm tracks targets by simply comparing pixels colors in each frame to prior known colors of the target of interest. However this method does not account for other types of underwater distortions such as particles, depth, adsorption and refraction, and will fail if multiple object of the same color such as fishes or turtles, are present in the frame. D. Walther et al [28] proposed a system for tracking multiple objects in underwater environment. It uses selective attention algorithm to reduce the complexity of muti-target tracking. Detection of new objects is done using saliency-based bottom up attention system, and kalman filter is used to track the centroid of detected objects. However, the speed and accuracy still have a lot of room for improvement.

D. Kim et al [27], proposed using weighted correlation coefficient for underwater target tracking. The approach consist of using texture and color based features to perform template matching of target under various lighting conditions. Objects are detected using multiple template based selection and tracking is done using mean shift based object tracking method. The algorithm also recourse to Gaussian smoothing and histogram equalization to compensate for distortion. The proposed method is robust to illumination change but does not perform well for other forms of underwater distortion.

In this paper, we demonstrate the weaknesses of existing state of the art tracking algorithms in underwater visual data and propose a dataset and benchmark to encourage development of more robust underwater tracking algorithms.

III. Dataset

As mentioned earlier, there currently exist numerous object tracking dataset and Benchmark. However, most of these dataset focus on open air or surface tracking.

978-1-7281-5092-5/19/$31.00 ©2019 IEEE

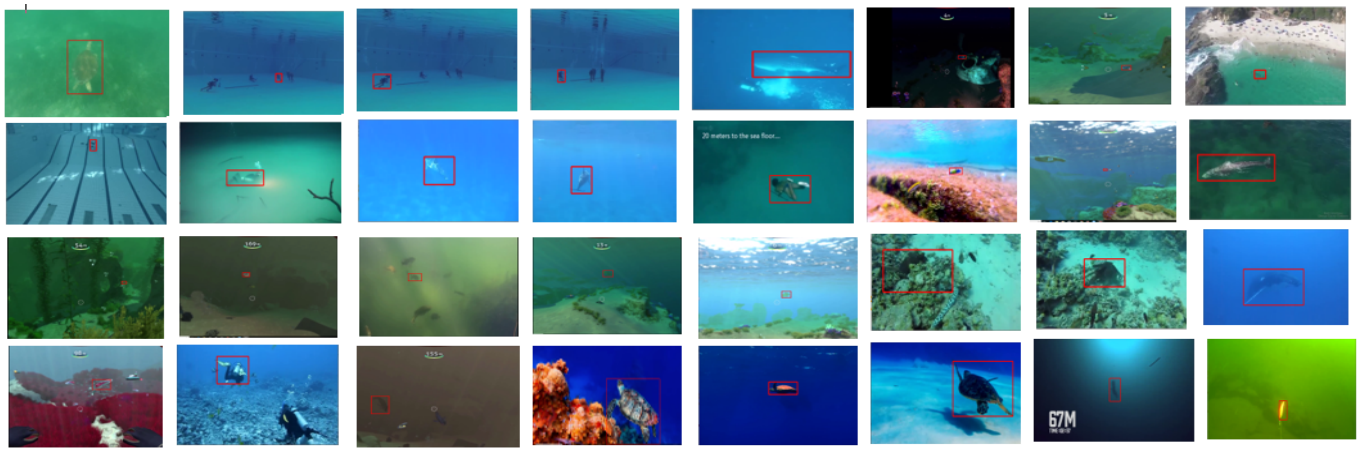

Figure 2. Sample tracking videos from our UOT32 dataset

Figure 2 shows sample frames from videos in the dataset with the ground truth bounding box for the target in the corresponding frame.

Our dataset consists of 32 videos for a total of 933 seconds and 24241 annotated frames. Each video has on average 29.15 seconds and 757.53 frames. Videos data are sourced from YouTube and consist of a combination of naturally distorted underwater visual data as well as artificial underwater data from Subnautica [30]. Figure 3 shows the distribution of the total number of frames per video in the dataset.

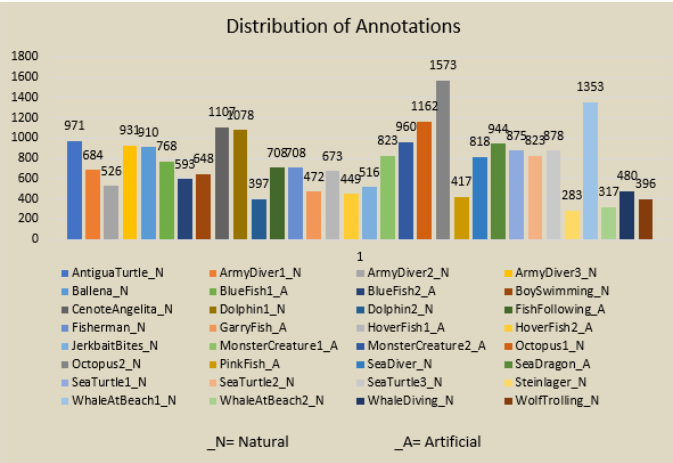

Figure 3. Distribution of Annotated frames

IV. Evaluation Methodology

For adequate and fair evaluation of existing state of the art object tracking algorithms on our dataset, we run tracking algorithms of up to 15 popular trackers whose original source code are made available to the public. Most of the trackers used here for evaluation are listed among the top performing tracking algorithms from the OTB and VOT object tracking challenge.

We adopt two common evaluation metrics to quantitatively analyse the performance of selected tracking algorithms on our UOT32 dataset. Just like in the OTB and VOT benchmark, we perform the One Pass Evaluation (OPE) and measure the precision as well as the success rate of each trackers over the entire 32 videos dataset. Precision is measured as the euclidean distance measured in pixels, between the center of the ground truth bounding box $C^{grd-bbox}$ and the tracker bounding box $C^{tkr-bbox}$. A threshold of up to 50 pixels is used for ranking trackers performance and plotting the corresponding precision [31].

$$P = \sqrt{C^{grd-bbox^2} - C^{tkr-bbox^2}}$$

Success rate on the Other hand measures the amount of overlap or Intersection over Union (IoU) of pixels between the ground truth and tracker bounding boxes. The performance on a sequence is measured by counting the total number of frames in which the bounding boxes overlap are greater than a given threshold value. The tracking algorithms are ranked using the Area Under Curve (AUC).

$$S = \frac{R^{tr} \bigcap R^{gt}}{R^{tr} \bigcup R^{gt}}$$

Where R^{tr} and R^{gt} represent the tracker and ground truth bounding box for a given frame respectively.

978-1-7281-5092-5/19/$31.00 ©2019 IEEE

V. Tracking Benchmark

The evaluation is conducted on an Alienware Area-51M 2080 Laptop with configurations as follows: 9th Generation Intel® Core™ i9-9900 (8-Core, 16MB Cache, up to 5.0GHz w/ Turbo Boost); NVIDIA® GeForce RTX™ 2080 11GB GPU; 64GB (2x32GB) RAM, DDR4, 2666MHz. For the One Pass Evaluation, each tracker is tested on all 32 videos and 24241 frames.

Overall Performance

The plot in Figure 4 shows the average Frame Per Second (FPS) of each tracker across all videos in the dataset. To evaluate the average FPS performance of a particular tracker, we first measure the average FPS of the tracker on each individual sequence.

The Overall FPS performance of of each tracker is then computed as the mean of average FPS for all sequences. In a similar fashion, we compute the Overall Precision as well as the Overall Success Rate performance of each tracker on the overall dataset, all in One Pass Evaluation fashion.

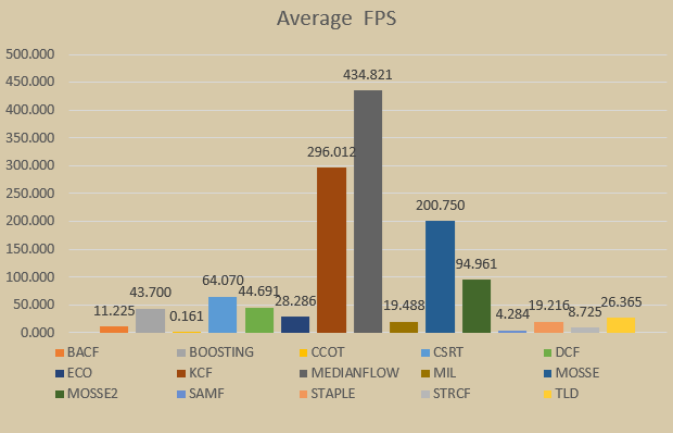

Figure 4. Average FPS Stats

As can be seen on the plots in Figure 4, 5 and 6, C-COT tracker has the highest precision and success at **0.472** and **0.397** respectively, but at an extremely slow overall tracking speed at **0.161 FPS** which is way below 1 FPS and not so useful. MEDIANFLOW on the flip side has the highest tracking speed at **434.821 FPS** but with a lower precision and success rate at **0.162** and **0.197** respectively. CSRT,BOOSTING, KCF, MOSSE, all perform at

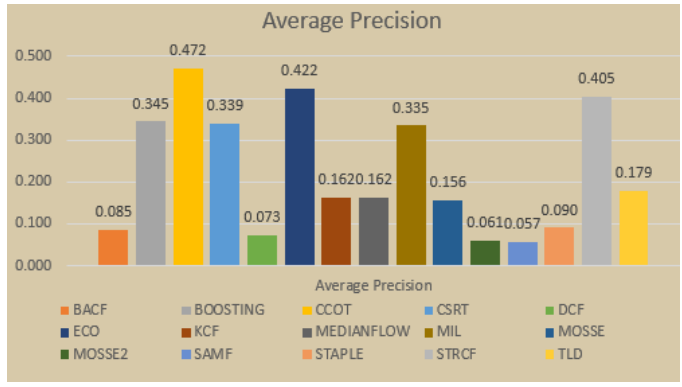

Figure 5. Average Precision Stats

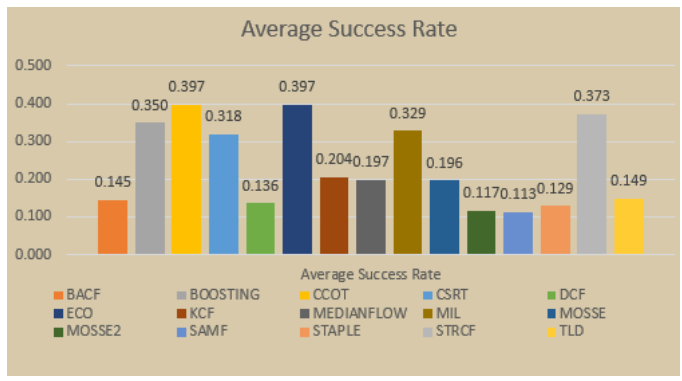

Figure 6. Average Success Rate Stats

real time speed **64, 43, 296, 200 FPS** respectively. Hoever, at reasonably low precision **0.339, 0.345, 0.162, 0.156** and sucess rate **0.318, 0.350, 0.204, 0.196** respectively. Figure 7 depicts the One Pass Evaluation of the Precision and Success Rate plots of the trackers.

Even with a pixel threshold of up to 50px, the precision of these state of the art trackers don't come close to their performance on open air dataset and benchmarks. Table I shows how some top performing trackers of the OTB50 and OTB100 benchmarks, perform on our Underwater Object Tracking (UOT) dataset.

This table substantiates our initial claim that the performance of existing state of the art algorithms considerably degrade when used in tracking objects in underwater environment as a result of the numerous aforementioned distortions. The results in Table I reveals that CCOT, ECO and STRCF trackers still rank as the top performing trackers on our UOT dataset, but with a much degraded performance.

978-1-7281-5092-5/19/$31.00 ©2019 IEEE

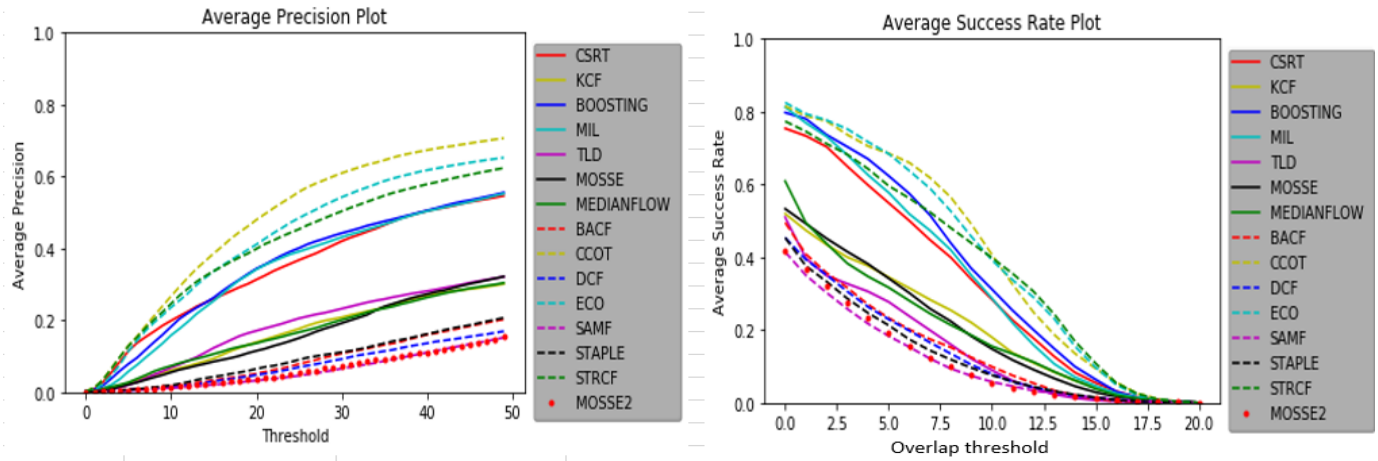

Figure 7. OPE Precision and Success Rate

Object Tracking Dataset						
Tracker	OTB50 precision	OTB50 success	100 precision	OTB100 success	**UOT32 Precision**	**UOT32 Success**
ECO [19]	0.874	0.643	0.910	0.691	**0.422**	**0.397**
CCOT [20]	0.843	0.614	0.898	0.671	**0.472**	**0.397**
STAPLE [21]	0.681	0.509	0.784	0.581	**0.090**	**0.129**
SAMF [25]	0.650	0.469	0.751	0.553	**0.057**	**0.113**
KCF [8]	0.611	0.403	0.696	0.477	**0.162**	**0.204**
BACF [23]			0.642		**0.085**	**0.145**
DCF [24]			0.733	0.598	**0.073**	**0.136**
STRCF [22]					**0.405**	**0.373**

Table I
PERFORMANCE COMPARISON ON UNDER WATER DATASET

VI. CONCLUSION

In this paper, we investigated the performance of current state of the art tracking algorithm in underwater environment in which visual data is distorted by refraction, reflection, particles, light, depth and much more; as opposed to open air object tracking which has been the principal focus of prior object tracking benchmarks. We introduce a rich and diverse underwater dataset with up to 32 videos and a total of 24241 annotated frames, from various distorted underwater environment. The results of our analysis illustrated in the plots and table in the Result section above clearly shows that top performing object tracking algorithms evaluated on open air environment, considerably degrade in performance when evaluated in underwater environment due to the inherent distortions present in underwater visual data. Hopefully, the UOT32 dataset presented in this paper will serve as a benchmark for developing new tracking algorithm more suited and more robust for underwater target tracking purposes.

978-1-7281-5092-5/19/$31.00 ©2019 IEEE

REFERENCES

[1] M. Shah A. Yilmaz, O. Javed, "Object tracking: A survey," *ACM Computing Surveys (CSUR)*, vol. 38, no. 13, April 2006.

[2] A. Berg, J. Ahlberg, and M. Felsberg, "A thermal object tracking benchmark," in *2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Aug 2015, pp. 1–6.

[3] Matej Kristan, Jiri Matas, Ales Leonardis, Michael Felsberg, Luka Cehovin, Gustavo Fernandez, Tomas Vojir, Gustav Hager, Georg Nebehay, and Roman Pflugfelder, "The visual object tracking vot2016 challenge results," in *Springer (Oct 2016)*, October 2016.

[4] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang, "Online object tracking: A benchmark," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013.

[5] Anton Milan, Laura Leal-Taixé, Ian D. Reid, Stefan Roth, and Konrad Schindler, "MOT16: A benchmark for multi-object tracking," *CoRR*, vol. abs/1603.00831, 2016.

[6] A. Li, M. Lin, Y. Wu, M. Yang, and S. Yan, "Nus-pro: A new visual tracking challenge," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 335–349, Feb 2016.

[7] Michael Felsberg, Amanda Berg, Gustav Hager, Jorgen Ahlberg, Matej Kristan, Jiri Matas, Ales Leonardis, Luka Cehovin, Gustavo Fernandez, Tomas Vojir, Georg Nebehay, and Roman Pflugfelder, "The thermal infrared visual object tracking vot-tir2015 challenge results," in *The IEEE International Conference on Computer Vision (ICCV) Workshops*, December 2015.

[8] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, March 2015.

[9] Shu-Kang Tu Shiuh-Ku Weng, Chung-Ming Kuo, "Video object tracking using adaptive kalman filter," in *Journal of Visual Communication and Image Representation*, December 2006, vol. 17, pp. 1190–1208.

[10] Mengdan Zhang, Junliang Xing, Jin Gao, Xinchu Shi, Qiang Wang, and Weiming Hu, "Joint scale-spatial correlation tracking with adaptive rotation estimation," in *The IEEE International Conference on Computer Vision (ICCV) Workshops*, December 2015.

[11] V. Buddubariki, S. G. Tulluri, and S. Mukherjee, "Multiple object tracking by improved klt tracker over surf features," in *2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, Dec 2015, pp. 1–4.

[12] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1409–1422, July 2012.

[13] H. Grabner, M. Grabner, and H. Bischof, "Real-time tracking via on-line boosting," in *Proc. BMVC*, 2006, pp. 6.1–6.10, doi:10.5244/C.20.6.

[14] Alan Lukežič, Tomáš Vojíř, Luka Čehovin Zajc, Jiří Matas, and Matej Kristan, "Discriminative correlation filter tracker with channel and spatial reliability," *International Journal of Computer Vision*, vol. 126, no. 7, pp. 671–688, Jul 2018.

[15] B. Babenko, Ming-Hsuan Yang, and S. Belongie, "Visual Tracking with Online Multiple Instance Learning," in *CVPR*, 2009.

[16] David Held, Sebastian Thrun, and Silvio Savarese, "Learning to track at 100 fps with deep regression networks," in *Computer Vision – ECCV 2016*, Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, Eds., Cham, 2016, pp. 749–765, Springer International Publishing.

[17] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2010, pp. 2544–2550.

[18] Z. Kalal, K. Mikolajczyk, and J. Matas, "Forward-backward error: Automatic detection of tracking failures," in *2010 20th International Conference on Pattern Recognition*, Aug 2010, pp. 2756–2759.

[19] Fahad Shahbaz Khan-Michael Felsberg Martin Danelljan, Goutam Bhat, "Eco: Efficient convolution operators for tracking," in *CVPR*, 2017.

[20] Fahad Khan Michael Felsberg Martin Danelljan, Andreas Robinson, "Beyond correlation filters: Learning continuous convolution operators for visual tracking," in *ECCV*, 2016.

[21] Stuart Golodetz Ondrej Miksik Philip H.S. Torr Luca Bertinetto, Jack Valmadre, "Staple: Complementary learners for real-time tracking," in *CVPR*, 2016.

[22] Wangmeng Zuo Lei Zhang Ming-Hsuan Yang Feng Li, Cheng Tian, "Learning spatial-temporal regularized correlation filters for visual tracking," in *CVPR*, 2018.

[23] Simon Lucey Hamed Kiani Galoogahi, Ashton Fagg, "Learning background-aware correlation filters for visual tracking," in *ICCV*, 2017.

[24] Fahad Shahbaz Khan Michael Felsberg Susanna Gladh, Martin Danelljan, "Deep motion features for visual tracking," in *ICPR*, 2016.

[25] Jianke Zhu Yang Li, "A scale adaptive kernel correlation filter tracker with feature integration," in *ECCV workshop*, 2014.

[26] D. Angela, C. Ion, I. Cornel, B. Diana, and P. Teodor, "Underwater object tracking using time frequency signatures of acoustic signals," in *OCEANS 2014 - TAIPEI*, April 2014, pp. 1–5.

[27] D. Kim, D. Lee, H. Myung, and H. Choi, "Object detection and tracking for autonomous underwater robots using weighted template matching," in *2012 Oceans - Yeosu*, May 2012, pp. 1–5.

[28] D. Walther, D. R. Edgington, and C. Koch, "Detection and tracking of objects in underwater video," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, June 2004, vol. 1, pp. I–I.

[29] Stéphane Bazeille, Isabelle Quidu, and Luc Jaulin, "Color-based underwater object recognition using water light attenuation," *Intelligent Service Robotics*, vol. 5, no. 2, pp. 109–118, Apr 2012.

[30] Subnautica, "https://unknownworlds.com/subnautica/," .

[31] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang, "Object tracking benchmark," in *IEEE Transactions on Pattern Analysis and Machine Intelligence 37(9), 1834–1848)*, 2015.

978-1-7281-5092-5/19/$31.00 ©2019 IEEE