Multimed Tools Appl (2017) 76:23117–23138
DOI 10.1007/s11042-017-4833-4

Marine animal classification using UMSLI in HBOI optical test facility

Zheng Cao[1] · José C. Príncipe[1] · Bing Ouyang[2] · Fraser Dalgleish[2] · Anni Vuorenkoski[2] · Brian Ramos[2] · Gabriel Alsenas[2]

Received: 31 July 2016 / Revised: 9 May 2017 / Accepted: 17 May 2017/
Published online: 24 May 2017

Abstract Environmental monitoring is a critical aspect of marine renewable energy project success. A new system called Unobtrusive Multistatic Serial LiDAR Imager (UMSLI) has been prepared to capture and classify marine life interaction with electrical generation equipment. We present both hardware and software innovations of the UMSLI system. Underwater marine animal imagery has been captured for the first time using red laser diode serial LiDAR, which has advantages over conventional optical cameras in many areas. Moreover, given the scarcity of existing underwater LiDAR data, a shape matching based classification algorithm is proposed which requires few training data. On top of applying shape descriptors, the algorithm also adopts information theoretical learning based affine shape registration, improving point correspondences found by shape descriptors as well as the final similarity measure. Within Florida Atlantic University's Harbor Branch

✉ Zheng Cao
zcao87@ufl.edu

José C. Príncipe
principe@cnel.ufl.edu

Bing Ouyang
bouyang@fau.edu

Fraser Dalgleish
fdalglei@fau.edu

Anni Vuorenkoski
adalglei@fau.edu

Brian Ramos
bramos5@fau.edu

Gabriel Alsenas
galsenas@fau.edu

[1] Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA

[2] Harbor Branch Oceanographic Institute, Florida Atlantic University, Boca Raton, FL, USA

Oceanographic Institute optical test facility, experimental LiDAR data are collected through the front end of the UMSLI prototype, on which the classification algorithm is validated.

Keywords UMSLI · Optical test facility · Shape matching · Similarity measure

1 Introduction

Marine hydrokinetic (MHK) projects are composed of undersea power generating equipment that converts the energy of waves, tides, or ocean currents into electricity. An imperative objective when deploying MHK devices is gaining an understanding of potential harmful interactions between marine animals and equipment at proposed development sites. Therefore, a high priority regulatory expectation exists to observe marine life interaction with such devices.

Underwater video observation of MHK scenes is typically accomplished with acoustical or optical cameras [24]. Available active acoustics technologies for fisheries monitoring can be categorized as either acoustic cameras or sounders. Active acoustic cameras, such as Dual Frequency Identification Sonar (DIDSON) [19] use acoustic lens technology, which forms images with greater detail than found in conventional sonars. While reasonably good estimates of feature dimensions can be extracted from DIDSON images, that level of detail is only available at relatively short ranges and with an angular field of view where image resolution is highest. Consequently, data bandwidth and storage become problematic. 2D imaging sonars can also provide high-quality output, but with limited angular field of view and range [6]. Acoustic sounders, on the other hand, are a standard instrument of fisheries hydroacoustics. These instruments can reach a long distance (hundreds of meters), but lack resolution and evaluation intuitiveness. Unlike active acoustic solutions, the primary advantage of using optical approaches is high resolution contrasted scene descriptions essential for detailed observations and object classification. Another advantage of optical imagery over acoustic data, is that it is more intuitive and easier to understand for humans. However, optical cameras are most effective when significant ambient light is present and when turbidity is low. Even the most sophisticated commercially-available underwater camera technologies require significant artificial white light to illuminate low light scenes. This is not desirable for MHK monitoring, because artificial light can alter the behavior of the animals being monitored. For example, it has been observed that marine life is attracted to light-emitting sources of wavelengths within their visual light frequency range.

The red laser diode serial Light Detection and Ranging (LiDAR) imager system we have developed combines the advantages of traditional optical and acoustic solutions while overcoming their disadvantages when applied to MHK environmental monitoring. This new approach is an adaptation of an existing technology that can be easily mounted on or around different classes of MHK equipment. The system is compact, cost-effective, and uses red laser illumination to be invisible and eye-safe to marine animals [20]. The equipment is designed for long-term, maintenance-free operations. It generates a sparse primary dataset that only includes detected anomalies, thus allowing for efficient, real-time, automated, low bandwidth animal classification/identification. The system is known as Unobtrusive Multistatic Serial LiDAR Imager (UMSLI). Applying LiDAR to MHK environmental monitoring is advantageous in several aspects. Firstly, red laser is beyond the visible wavelength range to most marine life, making it unobtrusive. Red laser illuminators can also be configured below the maximum permissible exposure (MPE) limit for humans. Second, unlike conventional optical camera whose focus is governed by the lens, LiDAR imagery focus is

governed by the laser beam spatial characteristics. LiDAR imagery thus remains in focus throughout the entire range, which is ultimately limited by attenuation at approximately 5m to 8m in typical coastal conditions [14]. Third, the system detection limit and Signal-to-Noise-Ratio (SNR) are superior to cameras, due to the rejection of both backscattering and ambient solar background. This is achieved by time-gating, as well as more concentrated photon energy using a single element detector [7, 8, 23]. Fourth, the fact that laser beams have single frequency makes them strictly parallel inside seawater, enabling the measurement of an object's size.

One critical part of UMSLI software design is the classifier, which labels an incoming instance that has been detected. Compared to optical images of marine animals in a previous study [3], marine animal LiDAR imagery is unique because there are virtually no existing open-source underwater animal life LiDAR images available. As such, the popular convolutional neural network (CNN) approach is not a good choice for UMSLI, as CNN needs to be trained from a large number of labeled data. Pretrained CNN using ImageNet [10] is not valid either, because natural images and LiDAR images are so different that domain adaptation will be problematic. This paper proposes a novel classifier based on shape matching. LiDAR image of an object (e.g. a fish) is 2-D grayscale after time integration. The shape of the object's boundary is highly descriptive of the object, which is the reason that a shape-based approach is used. Most shape matching techniques [1, 11, 15, 27] develop a descriptor for every point on a shape, then compare a pair of shapes by matching the descriptors and calculating the matching cost. This process puts two shapes in a high dimensional space, yet it overlooks the possibility of comparing the shapes directly in the original 2-D space. This paper goes one step further by utilizing the descriptors for affine registration and its accompanying similarity measure, which is inspired by concepts from information theoretical learning [13, 16]. Such registration will in turn yield more accurate point correspondences as compared to using shape descriptors. Both the 2-D space similarity score and improved point correspondences will positively impact the final classification performance. It is worth noting that from a pure shape registration perspective, many non-rigid shape matching algorithms [17, 18] can achieve good results. However, non-rigid models are problematic for classification tasks, as excessive non-rigid morphing can cause objects of different species become indistinguishable. In addition, non-rigid models have a much larger computation expense. Therefore, an affine model is chosen over a non-rigid one. The classifier developed in this paper is well-suited for MHK site monitoring because of its fastness, accuracy and ease of implementation.

The UMSLI prototype was developed and validated at the optical test facility at Harbor Branch Oceanographic Institute (HBOI). The main focus of this paper will be the sensing and image understanding blocks of the prototype. The remainder of this paper is organized as follows. Section 2 introduces the UMSLI system (including the sensing hardware and other functionalities), as well as the experimental environment - the HBOI optical test facility. Section 3 details the shape matching based classifier. Section 4 discusses the results from the test tank experimental data. Section 5 concludes the paper.

2 Underwater LiDAR sensing system

The UMSLI is an integrated system with a two-tiered design (see Fig. 1) consisting of sensing hardware, image understanding (enhancement, detection, classification) and data archiving functional blocks. Level-one monitoring (i.e. focus of attention) detects the presence of any potential marine species to be monitored using a sparse scan. If detected, the

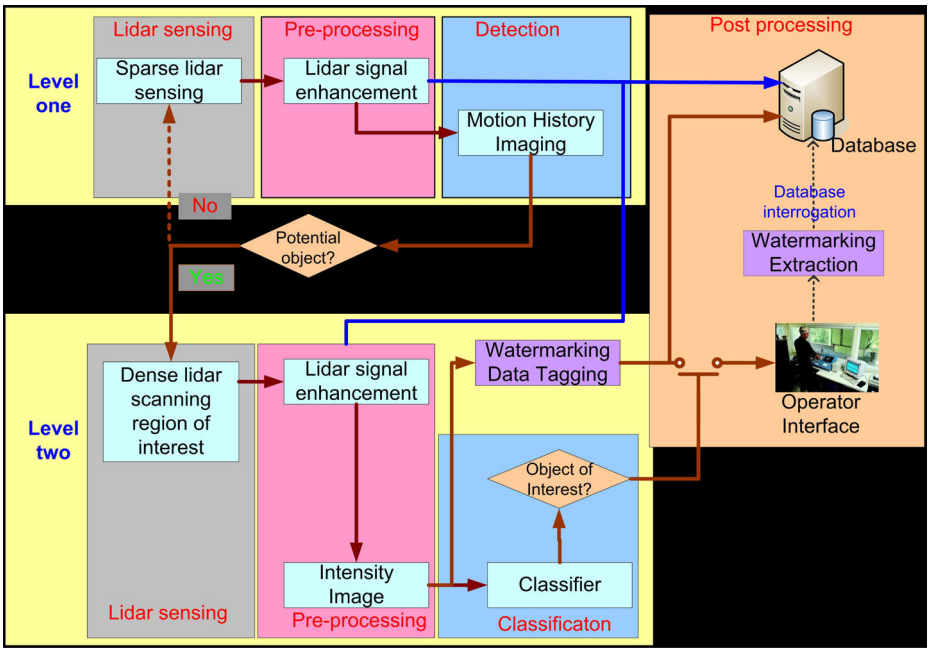


Fig. 1 Two-tiered processing framework of UMSLI

region of interest will be provided to the sensing front end to acquire classification-quality images via high resolution dense scan. Then classification is accomplished in Level-two classification block. The two-tiered detection/classification design enables real-time monitoring while significantly reducing the data volume for MHK site surveillance when the occurrences of the objects-of-interest are sparse. This design results in significant savings for the time and resource needed to deploy such monitoring system at an MHK site.

It should be noted that detection needs to be real-time, while classification does not. This is because real-time detection ensures that high resolution scan can be properly engaged by changing the laser focus area to the object of interest. On the other hand, certain latency for the classifier is allowed, since it takes time for an animal to possibly come in contact with the turbine after its first entrance into the field of view. As the animal is detected at 5-10m distance, the estimated time is 5-10 seconds assuming that the animal moves at 1m/sec.

2.1 Functional blocks

This section (Section 2.1) mainly explains the pre-processing, detection and post-processing functional blocks. The LiDAR sensing functional block is covered in the hardware description section (Section 2.2). The classification functional block will be detailed in Section 3.

2.1.1 Pre-processing

To enhance the signal-to-noise ratio of the LIDAR traces, the signal conditioning sub-blocks (see Fig. 2a) prior to level-one detection and level-two classification adopts the technique developed in [22]. The LiDAR returns will be enhanced prior to any further analysis. The

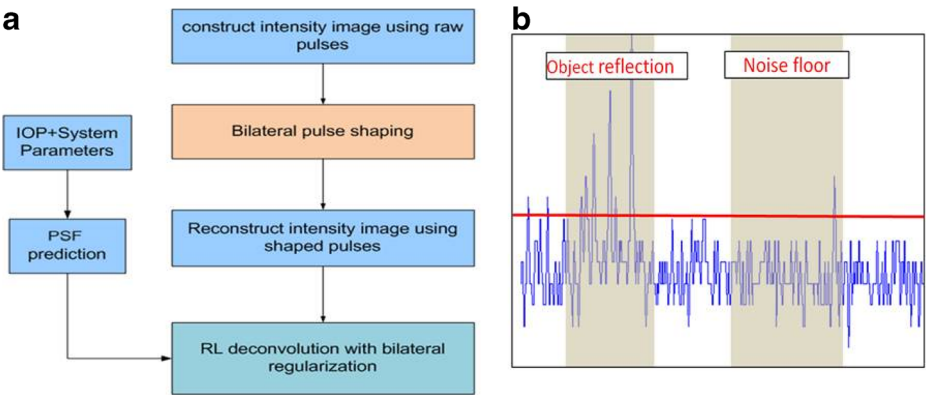


Fig. 2 **a** Processing flow of bilateral pulse shaping LiDAR signal enhancement; (**b**) Illustration of the noise floor and low SNR object reflection in a LiDAR trace

noise floor will be established through historical monitoring, as shown in Fig. 2b. The spatial correlation of the adjacent LiDAR pulses is exploited via the bilateral principle, followed by the deconvolution using the point spread function (PSF). The PSF can be predicted by a high-fidelity radiative transfer model: Electro-Optic DEtection Simulator (EODES) developed jointly by HBOI and Metron Inc. [12].

2.1.2 Detection and segmentation

Object detection is the foremost step for classification. A simple detector will threshold the image. Groups of adjacent pixels with gray-level values higher than the threshold form "blobs". If any blob contains a given threshold number of pixels (e.g. 100), it will be considered as an object. A valid threshold value choice is $median(image) - 10$. Since input to detection is the low resolution sparse scan image, motion history imaging (MHI) can be performed to increase the "footprint" of a moving object, typically an animal. MHI captures the signature of a moving object by computing the binary frame difference at two time instances. An illustration of the MHI of a fast moving object (e.g. a dolphin) is shown in Fig. 3. If the object is static or slow moving, the MHI step can be bypassed to reduce the reaction time during the detection stage. Note that detection locates the objects but does not give exact segmentation for the objects.

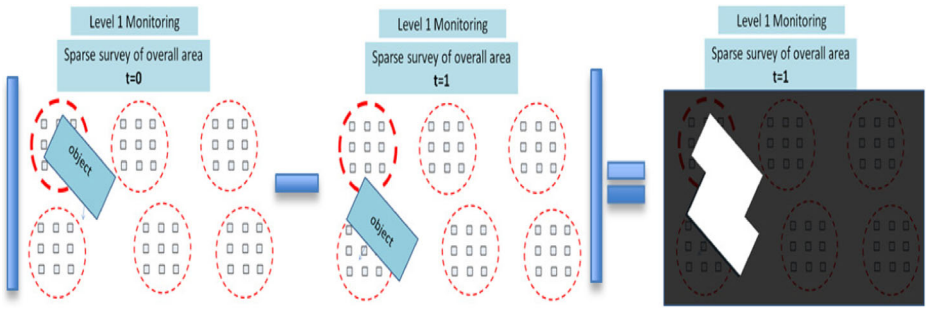


Fig. 3 MHI of fast moving object

If an object of potential interest is detected, it must be consequently converted into contour which is the input to the classifier. Following the detection step, a bounding box is put onto every object. The images are then processed such that the contrast is enhanced and the pixel value is constrained between 0 and 1:

$$I(x, y) = 1 - \exp\left(\frac{256 - I_0(x, y)}{\sigma(256 - \min(I_0)}\right) \tag{1}$$

In (1), $\sigma = 0.4$. Automated segmentation for the processed images can be performed by GrabCut [25]. GrabCut is an iterative method that minimizes the energy function E, in order to separate background and foreground/object:

$$E(\boldsymbol{\alpha}, \mathbf{k}, \boldsymbol{\theta}, \mathbf{z}) = U(\boldsymbol{\alpha}, \mathbf{k}, \boldsymbol{\theta}, \mathbf{z}) + V(\boldsymbol{\alpha}, \mathbf{z}) \tag{2}$$

In (2), U and V stand for data term and smoothness term, respectively. The data term is

$$\begin{aligned} U(\boldsymbol{\alpha}, \mathbf{k}, \boldsymbol{\theta}, \mathbf{z}) = & \sum_n (-\log \pi(\alpha_n, k_n) + \tfrac{1}{2} \log |\sigma(\alpha_n, k_n)| \\ & + \tfrac{1}{2\sigma(\alpha_n, k_n)} [z_n - \mu(\alpha_n, k_n)]^2) \end{aligned} \tag{3}$$

and the smoothness term is

$$V(\boldsymbol{\alpha}, \mathbf{z}) = \sum_{(m,n) \in C} -\gamma [\alpha_m - \alpha_n] \exp(-\beta ||z_m - z_n||) \tag{4}$$

In both terms, α specifies whether any pixel z_n belongs to the background ($\alpha_n = 0$) or foreground ($\alpha_n = 1$). In term U, k ranges from 1 to K, K being the number of Gaussian Mixtures (described by $\boldsymbol{\theta}$ which includes coefficient π, mean μ and variance σ) in either background or foreground. Term U can be viewed as the sum of the Mahalanobis distance between a pixel and its nearest Gaussian among all 2*K Gaussian mixtures. In term V, C indicates that m and n are neighboring pixels. Unlike term U, term V stresses the importance of putting neighboring pixels of similar intensity in the same background/foreground. The formulas (3) and (4) slightly differ from the original ones [25] in that the pixels z_n here are scalar values rather than 3*1 RGB vectors.

2.1.3 Post-processing

Data post-processing includes database index embedding and database index retrieval and query, both of which are based on digital watermarking. There are two possible watermark embedding approaches. The first is to replace the least significant bit (usually red channel) with watermark, the second is to use Zernike moments based embedding [21]. Comparatively, least significant bit replacement is simple to implement and has higher watermark embedding capacity, but it is more susceptible to high noise than Zernike moments based embedding.

2.2 Hardware design

The hardware design is based on the distributed serial pulsed laser imaging system concept [8, 9].

The UMSLI sensing front end consists of receiver(s) (Rx), transmitter(s) (Tx), and a digital signal processor (Fig. 4). Transmitters artificially "illuminate" a volume of water around an MHK device and receivers collect time-resolved returns from the laser pulses. The Rx and Tx components are off-the-shelf, inexpensive and robust. The system thus becomes

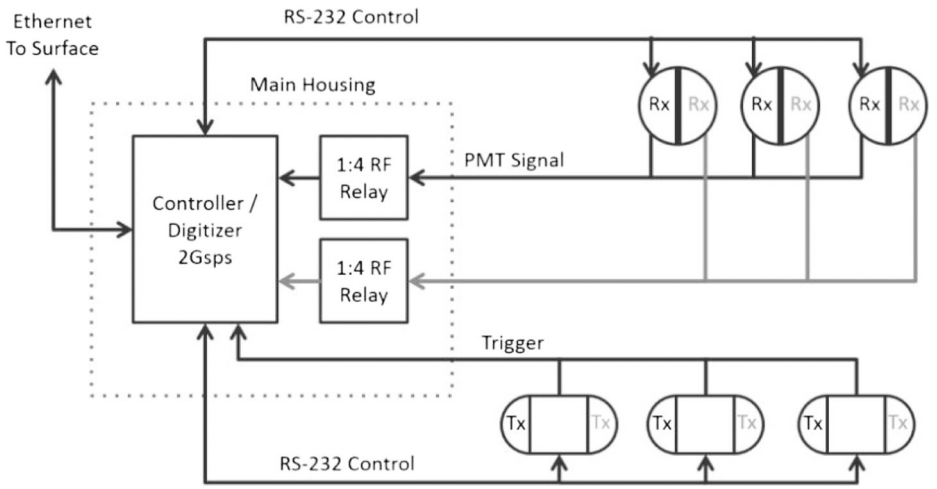


Fig. 4 UMSLI sensing front end

highly feasible with respect to Technology and Manufacturing Readiness Levels (TRL and MRL) and end-user operating costs. The initial prototype for each bidirectional transmitter housing Bill of Materials (BOM) cost is approximately $7k, while each bidirectional receiver BOM cost is $5k. The initial prototype system design has been validated in the HBOI optical test tank. The next phase of the development work will include the initial prototype system to be tested at-sea during the fall of 2016. This will allow for an evaluation of the performance of the sensor in real conditions at a range of depths and turbidities, and also allow for an assessment of the effectiveness of the proposed deployment methodology.

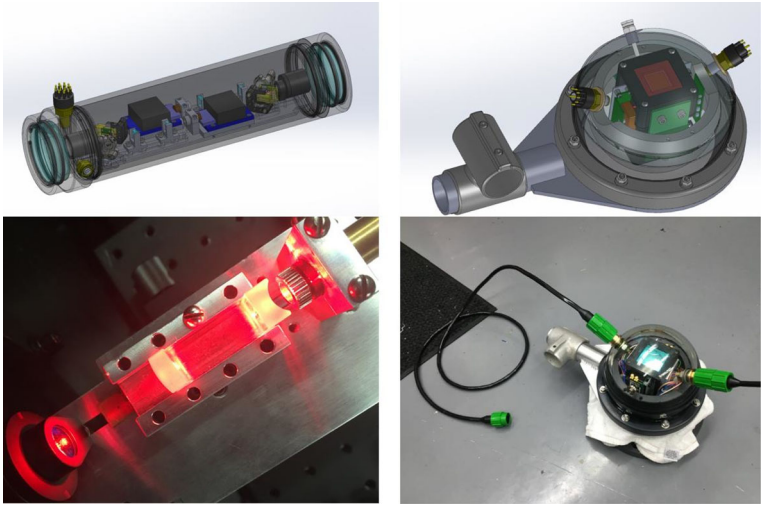


Fig. 5 Prototype bidirectional transmitter (*left*) and receiver (*right*) assemblies (CAD model – *top*; photos – *bottom*) which were used during the test tank validation and performance evaluation testing

The prototype bidirectional transmitter and receiver assemblies, which were used during the test tank testing phase, are shown in Fig. 5. The Tx contains a pair of bidirectional scanned LiDAR transmitters, each consisting of a high power laser diode at 638nm, driven by a pulse driver that can generate 10W peak power pulses of 5 ns pulse duration (FWHM) at up to 1 Mpps pulse repetition rate. The laser pulses are then scanned in a bidirectional raster scan pattern using an Analog Micromirror Device (AMD) over a total scan angle of 60 degrees by 60 degrees. The Rx consists of a 18cm diameter sphere containing a pair of bidirectional red sensitive, large aperture, high speed, high sensitivity receivers. Each receiver consists of a high pass filter to reject optical energy at wavelengths shorter than 630nm, a 24mm $\times$ 24mm Photomultiplier tube (PMT) module with 12% quantum efficiency at 638nm and a 2 ns FWHM impulse response time. The Tx and Rx housings are both connected to an electronics housing that contains a PXI command, control and acquisition computer with a maximum of four simultaneous 12 bit digitizer channels, with maximum sample rate of 2 Gsps. For the experiments described herein, the digitizer channel is 8 bits, operating at 1 Gsps.

UMSLI's hardware configuration grants many favorable properties for LiDAR sensing. For example, multiple wide-angle pulsed laser illuminators and an array of single element photon detectors can be configured to cover a 4pi steradian (i.e. omnidirectional) scene volume. Consequently, three dimensional features of animal targets can be retrieved and clutter is rejected from time-resolved information. Moreover, the transmitter operates in an adaptive mode and can be reconfigured very quickly. Under the sparse scan mode, it projects a sparse grid of pulses with higher peak power for longer range detection of animals within a zone of interest, which is greater than 10m in typical coastal conditions. Under the dense scan mode, it generates high resolution imagery for identification of marine life at closer range.

2.3 The HBOI optical test facility

The Ocean Visibility and Optics Lab at HBOI developed a unique test facility (Fig. 6) using the funding provided by the Office of Naval Research (ONR). The test facility is capable of extensive testing of many different electro-optical system configurations under a range of environmental conditions. During the validation tests within the test facility, the image quality is measured using a USAF-1952b technical target with 0.5m^2 dimensions, placed at 5m distance from the imaging system.

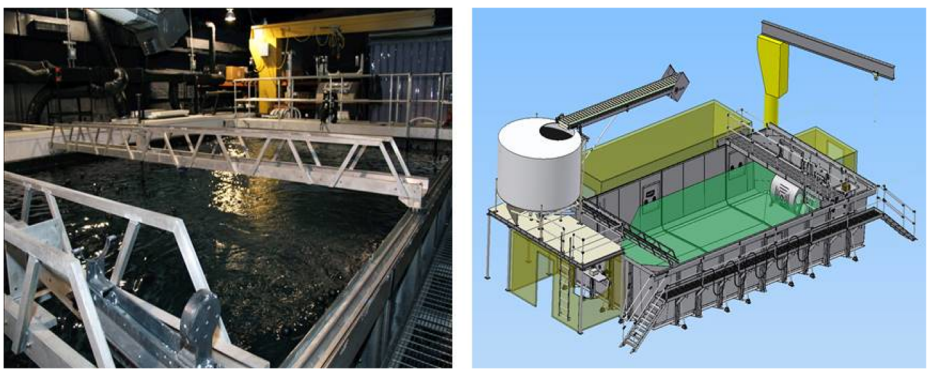


Fig. 6 HBOI underwater electro-optics testing facility

The facility contains a 40ft long, 22ft wide and 7ft deep test tank. There are three attached labs surrounding the test tank. Within these labs, there are in total seventeen different viewports allowing different system configurations to be investigated. There are flexible in-water target and illumination housing installation options, including a rotating target drum with a maximum stand-off distance of 38ft and a four-degree-of-freedom linear driver for in-water target and illuminator mounting. Circulation jets in the tank can maintain controlled particle suspensions (stable up to $c = 4m^{-1}$). Sensor carriages, underwater radiometers, AC-9 transmissometers (WET Labs) are installed to provide in-situ inherent optical properties measurements during an experiment. 70,000 black spheres can be deployed on the water surface to block out solar ambient light and minimize internal reflections during an experiment.

The image quality goal and actual outcomes are summarized in Fig. 7. In clear water, the image resolution is 0.25cm. At higher turbidity (attenuation coefficient $c = 0.73\ m^{-1}$), the achievable resolution is 1cm, which is above the original system design goal. Type images acquired during the tests in clear water and at a high turbidity of $c = 0.73\ m^{-1}$ (equivalent to turbid coastal condition with diver visibility of approximately 3 meters) are shown in Fig. 5.

For the marine animal classification experiment, taxidermy replicas of three marine species (amberjack, barracuda, and turtle) are mounted in on a four degrees-of-freedom robotic stage in the HBOI optical test tank. The experiment is detailed in Section 4.

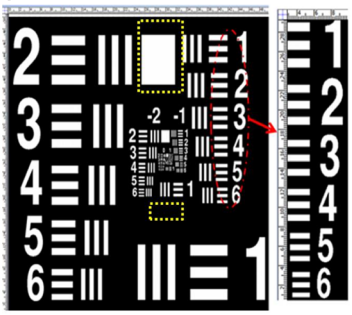
3 Shape matching-based classifier

In Section 3, a shape is characterized by N 2-D points on its (outer) contour. The word "shape" is used interchangeably with the point set. A "query shape" $\mathbf{Y}$ is the shape whose label is going to be found, whereas a "template shape" $\mathbf{X}$ has a known label.

3.1 Point correspondence

Point correspondence between a pair of shapes can be established in two steps. During the first step, the descriptor at every point on both shapes are calculated. Then in the second step, the best matches between the descriptors are found. SC [1] and IDSC [15] are among the best-known descriptors for measuring shape similarity. SC describes a point $\mathbf{x}_i$ by its

a



b

Turbidity level	CR	CSNR (raw image)	Smallest Resolvable Bar
Clear water GOAL	0.7	15	1cm
Clear water RESULTS	0.8	20	0.25cm
> 4 beam attenuation lengths GOAL	0.2	8	2cm
> 4 beam attenuation lengths RESULTS	0.3	3.23 (processed: 8.5)	1cm

Fig. 7 Image quality achieved with the UMSLI prototype. **a** Technical target used for evaluation; (**b**) Achieved image quality metric

Euclidean distance to all points $\mathbf{x}_j$ (j = 1,...,N) on the same shape, as well as the angle formed by the tangent line at $\mathbf{x}_j$ and the line connecting $\mathbf{x}_i$ and $\mathbf{x}_j$. With n_r bins for distance and n_θ bins for angles, a 2-D histogram is built. The histogram is then stretched to a $n_r * n_\theta$ vector $\mathbf{h}_{xi}$ which is the descriptor for $\mathbf{x}_i$. Conceptually, SC is a rich, global and rotational invariant descriptor. The IDSC is an extension of SC, which uses the shortest path inside the shape between two points for calculating "inner distance" and "inner angle", rather than using the line segment directly. IDSC performs better than SC on many shape discrimination tasks, especially those involving articulate invariant objects (e.g. scissors).

Following the calculation of descriptors, a cost matrix C will be built, whose $(i, j)^{th}$ entry denotes the cost of matching $\mathbf{y}_i$ to $\mathbf{x}_j$. Assume χ^2 distance is used:

$$C(\mathbf{y}_i, \mathbf{x}_j) = \sum_{k=1}^{n_r * n_\theta} \frac{[\mathbf{h}_{yi}(k) - \mathbf{h}_{xj}(k)]^2}{\mathbf{h}_{yi}(k) + \mathbf{h}_{xj}(k)} \tag{5}$$

The best correspondences $(\mathbf{y}_i, \mathbf{x}_{j(i)})$ (i = 1,...,N) will minimize the descriptor cost that is the sum of cost over all correspondence pairs:

$$descriptor\_cost = \sum_{i=1}^{N} C(\mathbf{y}_i, \mathbf{x}_{j(i)}) \tag{6}$$

The correspondences can be computed using the Hungarian algorithm [1], which is a combinatorial optimization algorithm that iteratively switches a matrix's rows such that the trace is minimized. For ease of use, the points of the template shape in Section 3.2 are reordered as $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$, where $\mathbf{x}_i$ corresponds to $\mathbf{y}_i$.

3.2 MCC affine registration

The principles of doing affine registration using the maximum correntropy criterion (MCC) have been presented [2] and is reviewed in this section. With the availability of point correspondence, affine registration can be formulated as an optimization problem. Affine transformation performed on template shape - ordered points $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$ can be expressed as $f(\mathbf{X}) = \mathbf{XA}$. $\mathbf{X}$ and transformation matrix $\mathbf{A}$ are originally 2-dimensional, but applying homogeneous coordinates makes them N*3 and 3*3, respectively. To find a reasonable $\mathbf{A}$, a common practice is to minimize the mean squared error (MSE) between transformed template $\mathbf{XA}$ and query $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^N$. However, MSE will become suboptimal in the presence of outliers (wrong correspondence). Instead, the maximum correntropy criterion (MCC) is chosen because of its robustness to outliers [16]:

$$\mathbf{A} = \text{argmax} \sum_{i=1}^{N} G_\sigma(\mathbf{x}_i\mathbf{A}, \mathbf{y}_i) \tag{7}$$

where $G_\sigma(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{||x-y||^2}{2\sigma^2})$. By taking derivative with respect to $\mathbf{A}$ and solving the equation one gets

$$\begin{aligned} \mathbf{D} &= \text{diag}(G_\sigma(f(\mathbf{x}_1), \mathbf{y}_1), ..., G_\sigma(f(\mathbf{x}_N), \mathbf{y}_N)) \\ \mathbf{A}_{new} &= (f(\mathbf{X})^T \mathbf{D} f(\mathbf{X}))^{-1}(f(\mathbf{X})\mathbf{DY}) \\ \mathbf{A}_j &= \mathbf{A}_{j-1}\mathbf{A}_{new}, \quad f(\mathbf{X}) = \mathbf{XA}_j \end{aligned} \tag{8}$$

Equation (8) is a fixed point solution whose convergence can be guaranteed [4]. $\mathbf{A}$ is initialized by $\mathbf{I}_{3*3}$. A maximum of 50 iterations are performed unless the stopping criterion $||\mathbf{XA}_j - \mathbf{XA}_{j-1}|| < 10^{-4}||\mathbf{Y}||$ is reached. The choice of kernel size σ is an important

Table 1 Summary of similarity measures used in the paper

Method 1	Method 2	Method 3	Method 4
SC cost	SC cost + instant corr. cost	IDSC cost	IDSC cost + DP corr. cost

Method 1 and 2 have been implemented in previous work [2]

issue. When σ becomes very large MCC (7) degenerates into MSE, which is not as good in dealing with outliers. Meanwhile, σ that is too small may cause the solution to stuck at local minima. A good initial kernel size σ is the mean distance of all points in the query d_m, which is obtained during the computation of shape descriptor. Annealing is necessary to ensure fast convergence. During the first 20 iterations, an annealing rate of 0.9 is used.

3.3 Shape similarity measure

Solving for point correspondence automatically gives the descriptor cost (6), which is the similarity measure used in literature [1, 15]. With the affine registration given in Section 3.2, it is expected that some new similarity measure that benefits from the registration may exist. One such measure is the "instant correntropy cost" [2]:

$$instantcorr\_cost(\mathbf{X}, \mathbf{Y}) = \sum_{i=1}^{N} G_\sigma(\mathbf{y}_i, f(\mathbf{x}_i)) \tag{9}$$

The instant correntropy cost has two advantages. First, although the majority of point correspondences are correct, many others are not (e.g. a fish's head matches with tail). The descriptor cost associated with these incorrect correspondences will be incorrect. On the other hand, because of correntropy's nice property of being local, the correntropy cost on the any incorrect correspondence is close to zero. Second, the correntropy cost does not require extra computation because it derives naturally from (7). The kernel size is chosen as the σ after the last iteration during the fixed point solution.

Another registration-based similarity measure is called the dynamic programming (DP) correntropy cost. Consider the optimization problem

$$\text{argmin } A_\pi \sum_{i=1}^{N} d(\mathbf{y}_i, f(\mathbf{x}_{\pi(i)})), \text{ s.t. } \pi(i) \leqslant \pi(j), \ \forall i \leqslant j \tag{10}$$

In (10), $\mathbf{Y}$ and $f(\mathbf{X})$ are the pair of registered shapes, while d denotes Euclidean distance. This is because an intuitive way for finding a correspondent point on the other registered shape is to search the nearest point. The constraint is based on the fact that all points are on the contours and are hence ordered. Therefore, DP [5] can be applied to solve the optimization problem. One parameter associated with DP is the penalty τ for leaving $\mathbf{y}_i$ unmatched ($\pi(i) = 0$), which is set as d_m in Section 3.2. After the correspondences $(i, \pi(i))$ are found, the DP correntropy cost can be calculated:

$$DP\_corr\_cost(\mathbf{X}, \mathbf{Y}) = \sum_{i=1, \pi(i)\neq 0}^{N} G_\sigma(\mathbf{y}_i, f(\mathbf{x}_{\pi(i)})) \tag{11}$$

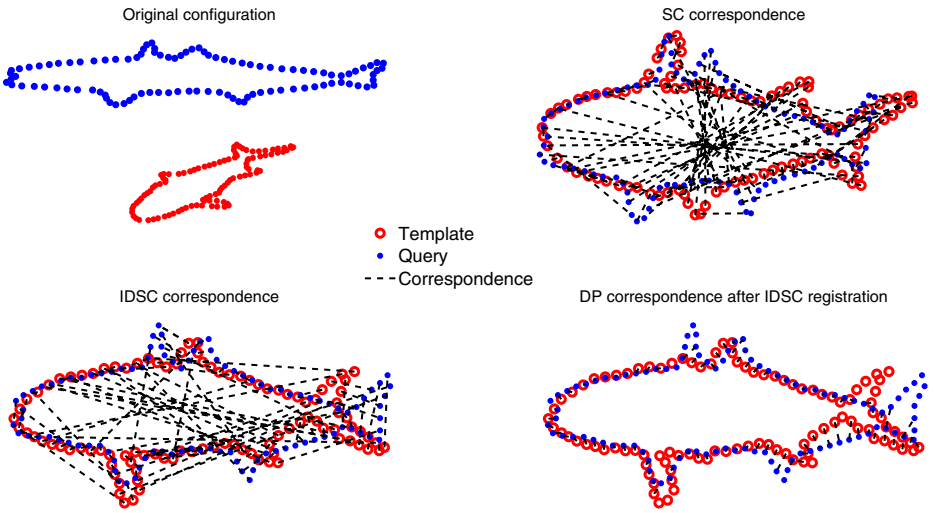


Fig. 8 Illustration of shape registration and similarity cost evaluation. The MCC affine registration results are good for both SC (*upper right*) and IDSC (*lower left*) descriptors. Yet neither descriptor actually gives perfect correspondences. The instant correntropy cost will "suppress" bad correspondences, while the DP correntropy cost is based on refined correspondence given by (10)

Similarity measures can mix with one another in a heuristic manner, such as linear combination [1]. A new similarity measures will thus be used in this paper:

$$new\_cost(\mathbf{X}, \mathbf{Y}) = \frac{descriptor\_cost(\mathbf{X}, \mathbf{Y})}{corr\_cost(\mathbf{X}, \mathbf{Y})} \tag{12}$$

In (11), $descriptor\_cost$ may denote SC cost or IDSC cost, and $corr\_cost$ may denote instant correntropy cost or DP correntropy cost. Table 1 summarizes the 4 similarity measures tested in this paper. The methods are further illustrated in Fig. 8.

The 4 methods are tested firstly in an artificial dataset named Kimia 99 [26]. The dataset consists of 9 different categories of objects, with 11 instances in each category. For any instance that serves as the query shape, 10 out of 98 matches that yield the smallest cost are found. The total number of top one to ten correct matches are presented in Table 2. Clearly, applying correntropy cost on top of descriptor cost renders better results than applying descriptor cost alone. For both this experiment and the experiment in Section 4, the parameters are set as $N = 100$, $n_r = 7$ and $n_\theta = 12$.

Table 2 Shape retrieval results for Kimia 99 dataset

Method	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
Method 1	99	97	97	97	96	94	94	88	84	81
Method 2	99	99	98	98	98	98	98	93	90	79
Method 3	99	97	96	95	90	85	82	76	66	50
Method 4	99	97	97	97	96	95	95	94	87	68

Method 2 outperforms Method 1, and Method 4 outperforms Method 3

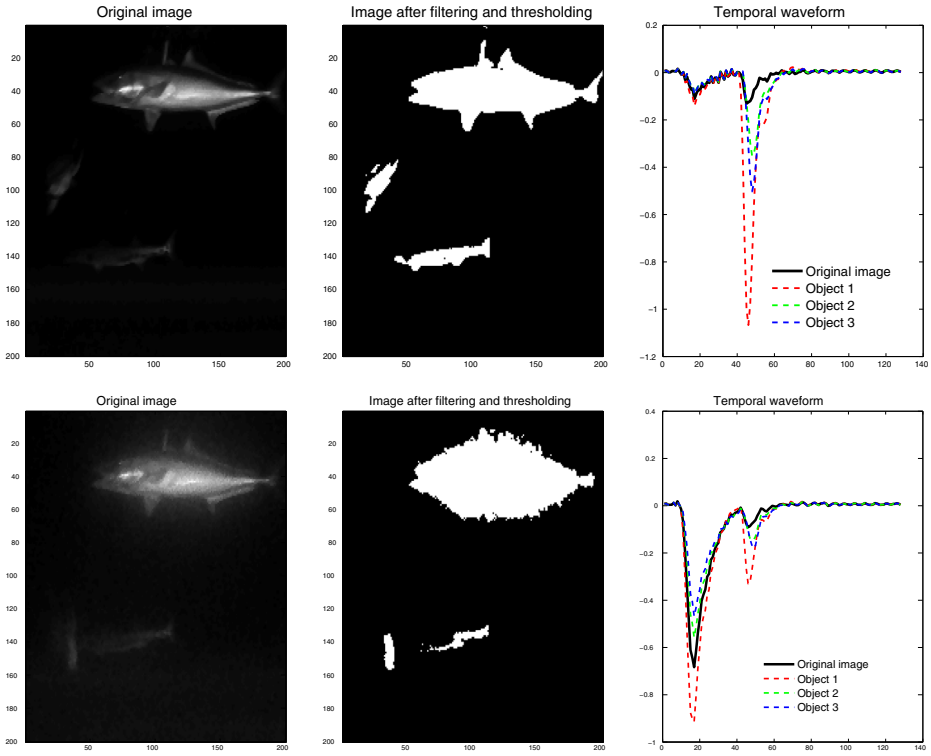


Fig. 9 *Left*: original LiDAR image (summed over time 1-128). *Center*: LiDAR image after bilateral filtering and thresholding. *Right*: temporal waveform on all image and on individual objects. The *top row* shows the clear water image while the *bottom row* shows $c = 0.73$ image

4 Discussion of experimental results

The experimental data are collected under two different turbidity levels: clear water and and at beam attenuation coefficient $c = 0.73$. For each turbidity level, 8 separate trials are taken place. Every trial produces an image that contains one amberjack, one barracuda, and one turtle, although one should consider neither the number nor specie of the animals to be known before classification. For each trial, the orientations of the objects are adjusted randomly before the images are acquired. The spatial resolution of an image is 200*200. If one looks from the temporal perspective, each pixel is a 1-D waveform of length 128.

4.1 Detection and contour extraction

The 1^{st} trial of clear water image and the 1^{st} trial of $c = 0.73$ image are taken as examples. The left plot of Fig. 9 shows the original image where pixels are summed over all 128 time instances. The solid black line in the right plot is the waveform summed over all pixels. Its first and second peaks are the backscatter and the LiDAR return of object(s) respectively, as backscatter typically precedes object LiDAR return. By time-gating to keep only the second peak (time 43–52), the objects become more prominent. It is seen that LiDAR return for

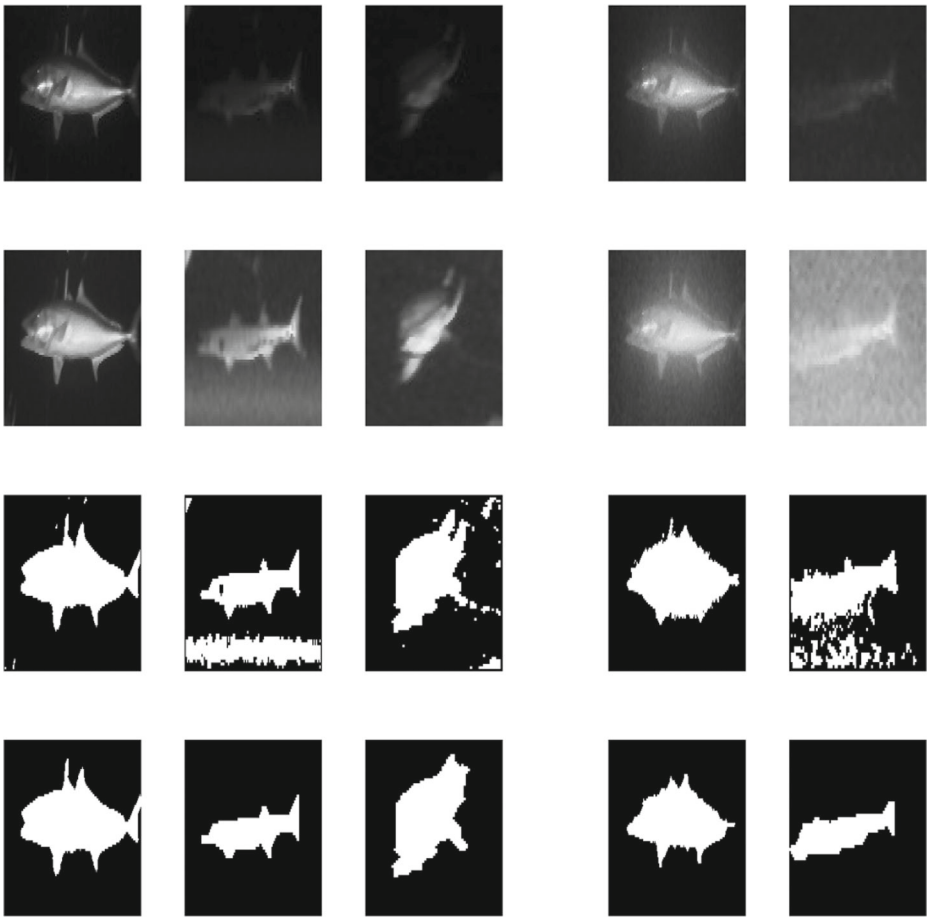


Fig. 10 From *top to bottom row*: objects with bounding box; objects after enhancement using (1); initialization of GrabCut; result of GrabCut. The first 3 columns are from a clear water image (*upper left* of Fig. 9), while the last 2 columns are from a turbid water ($c = 0.73$) image (*lower left* of Fig. 9)

$c = 0.73$ is much weaker than clear water image. The image then goes through bilateral filtering (Fig. 2a) as a pre-processing step and thresholding, resulting in the center plot of Fig. 9. It follows that 3 objects are identified in the image. The colored lines in the right plot depict the waveform of the 3 objects. As the peaks of the waveform heavily overlap with each other, it is impossible to separate the objects with temporal slicing. This means temporal information is the only information one can utilize here for detection and the following segmentation.

It should be noted that although there are 3 objects in every image, there are only 22 visually discernible objects in all 8 clear water images, and 16 discernible objects in images with $c = 0.73$. The indiscernible objects are either too small, too blurred or under the occlusion of another object.

For GrabCut segmentation, parameters are set as follows: $\mathrm{K} = 3$, $\beta = 1$, $\gamma = 100$ for all 8 clear water images. For the more turbid images with $c = 0.73$, γ is set to 1000. GrabCut needs an initialization of background and foreground, which is done by marking the top

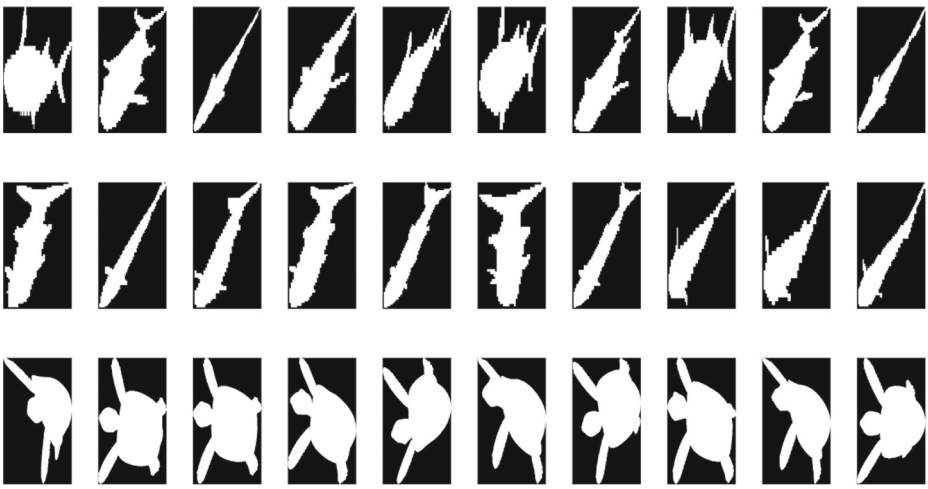


Fig. 11 Chosen template shapes. From *top to bottom*: amberjack, barracuda, turtle

25% pixels with largest intensity as foreground. The center row and bottom row of Fig. 10 show the initialization and final result of GrabCut.

4.2 Template generation

For each specie, a three-dimensional model is generated from a prototype model downloaded from the internet, using the software Blender. The 3-D model is then projected onto different 2-D planes, producing 256 2-D template shapes. As many of the templates share a considerable amount of similarity with others, it is desirable to choose a few "representatives" from all 256 templates, which will also significantly reduce computational time in the following classification step. This requires all templates to be represented in the form of vectors, on which a mode-seeking algorithm will be performed. To this end, one can build a 256*256 similarity matrix M using one of the aforementioned methods (e.g. method I). A shape is represented by the corresponding row vector in the similarity matrix after row-wise normalization. The K-means algorithm is then applied to the 256 vectors such that K clusters are found. Within each cluster, the "representative" is simply the vector that has the smallest L1-norm. For $K = 10$, the chosen templates are shown in Fig. 11.

4.3 Classification results

The query will be matched with all available templates, and classified as the specie of the template that yields the smallest cost (1 nearest neighbor classification).

The clear water query images has in total 8 amberjacks, 6 barracudas and 8 turtles. Three different usages of templates are considered. In the first case, all 256 simulated templates for each specie are used. In the second case, one randomly selects 10 templates out of all 256. For this case, 1000 Monte-Carlo trials are conducted and the average accuracy is reported. In the third case, 10 templates are chosen in the manner described in Section 4.2. All 4 methods for measuring similarity described in Section 3 are implemented, making it 12 combinations in total. For every template usage and every method, a 3*3 confusion matrix

Table 3 Classification result (FPR/FNR) of clear water images

	I	II	III	IV	Average of I-IV
All 256	0.21/0.39	0.14/0.26	0.09/0.17	0.09/0.17	0.13/0.25
Random 10	0.21/0.42	0.18/0.36	0.10/0.21	0.08/0.17	0.14/0.29
Chosen 10	0.16/0.32	0.12/0.24	0.07/0.14	0.04/0.10	0.10/0.20
Average of different template usages	0.19/0.37	0.14/0.29	0.08/0.17	0.07/0.14	

Methods I, II, III and IV refer to the methods mentioned in Section 3. Effectiveness of any method is best reflected in the last row (average of different template usages). Similarly, effectiveness of any template usage is best reflected in the last column

is obtained. The respective false positive rate (FPR) and false negative rate (FNR) values are calculated as follows. For any specie (i.e. amberjack, barracuda or turtle),

$$\begin{aligned} FPR &= (total\ false\ positives)\ /\ (total\ false\ positives + total\ true\ negatives) \\ FNR &= (total\ false\ negatives)\ /\ (total\ false\ negatives + total\ true\ positives) \end{aligned} \tag{13}$$

The FPR and FNR reported in Table 3 are the mean values of FPR and FNR for all 3 species.

For $c = 0.73$ query images (8 amberjacks, 3 barracudas and 5 turtles), the experiments for clear water images are repeated. In addition, it is critical to examine whether the query shapes generated from clear water images serving as templates can enhance classification accuracy. Therefore, two additional cases are added to the existing three. The first case uses the 10 chosen simulated templates altogether with 8 shapes from clear water images as templates, while only the 8 clear water image shapes are used as templates in the second case. The confusion matrices are shown in Table 4.

Result analysis A major conclusion to be drawn from Tables 3 and 4 is that Method II outperforms Method I (Method II results in smaller FPR and FNR), and Method IV outperforms Method III. In other words, applying shape matching always improves classification result. This is in accordance with the results on Kimia-99 data. Overall speaking, Method IV here also performs better than Method II. For the clear water case, the advantage of using 10 templates chosen by the method in Section 4.2 is obvious: not only does it top randomly chosen templates, but its performance is also even better than using all 256 templates. This indicates large redundancy in the original 256 templates. For $c = 0.73$, the incorporation of templates from clear water images is very beneficial, to the extent that the result of using

Table 4 Classification result (FPR/FNR) of c=0.73 images

	I	II	III	IV	Average of I-IV
All 256	0.31/0.64	0.09/0.22	0.13/0.23	0.09/0.15	0.16/0.31
Random 10	0.25/0.49	0.18/0.37	0.17/0.30	0.16/0.26	0.19/0.36
Chosen 10	0.23/0.46	0.16/0.35	0.17/0.28	0.14/0.23	0.18/0.33
Chosen 10 and clear water data	0.16/0.26	0.05/0.08	0.09/0.15	0.09/0.15	0.10/0.16
clear water data	0.16/0.26	0.05/0.08	0.09/0.15	0.07/0.11	0.09/0.15
Average of different template usages	0.22/0.42	0.11/0.22	0.13/0.22	0.11/0.18	

Table 5 Upper table: average time (in seconds) of computing the descriptor of one query

Method	image processing + GrabCut	compute SC	compute IDSC
Method 1,2	0.50	0.021	0
Method 3,4	0.50	0	0.033

Method	descriptor cost	registration	correntropy cost
Method 1	0.0069	0	0
Method 2	0.0069	0.0020	0.00027
Method 3	0.0069	0	0
Method 4	0.0069	0.0020	0.0011

Lower table: average time of computing pairwise similarity measure with one template

clear water images templates alone cannot be improved upon. This shows higher similarity between two sets of test tank data under different turbidity than between the simulation templates and test tank data.

Table 5 summarizes the time for classifying a query object. As templates can be stored, one needs only to get the query's contour and compute its descriptor, then calculate the pairwise similarities with all templates available. The upper table shows that image segmentation and post-processing is more costly than computing the descriptor for classification. The lower table suggests that the extra computational cost brought by shape registration and correntropy cost computation is relatively small compared to the descriptor cost itself.

5 Conclusion and future work

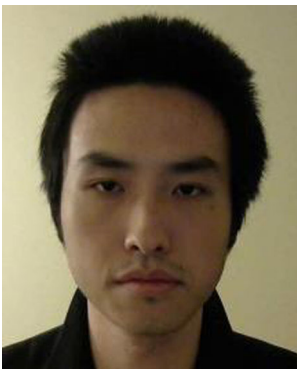
A UMSLI system with LiDAR sensing hardware, classification algorithm, and image understanding functional blocks has seen initial success in experiments conducted in a test facility. The incorporation of MCC affine registration into traditional shape descriptors boosts classification performance, while additional computation cost is low. It has also been confirmed that even without real LiDAR images, classification results are still satisfactory given that appropriate simulated data are used as templates, but results can be further improved if real data are available. To better integrate simulated and real data, manifold learning will be considered in the future for its ability to learn the space in which both real-world and simulated training data lie.

The upcoming field testing approach will involve attaching Rx, Tx, and digitizer housings to a deployment frame connected via electro-mechanical cable to a moored vessel. This will allow for baseline validation testing. Subsequent energetic site testing, if an MHK system is available, will involve strategically attaching the UMSLI components to the MHK device to demonstrate satisfactory field of view coverage. Although the system can consume low power (∼49W) and it is optimized to maximize data storage, long term deployments will require external power and data connection to shore. Further development and testing should include additional exposure to new marine life species in order to develop a comprehensively trained classifier.

Acknowledgments This work was supported in part by US Department of Energy contract DE-EE0006787 and FAU/HBOI internal fund.

References

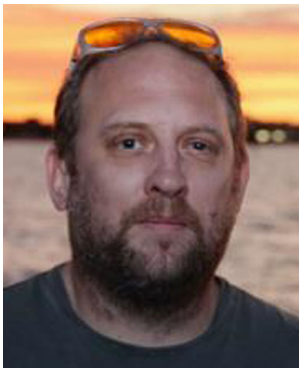
1. Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts. PAMI 24(24):509–522
2. Cao Z, Principe J, Ouyang B (2016) Information point set registration for shape recognition. In: Proc. ICASSP
3. Cao Z, Principe JC, Ouyang B, Dalgleish F, Vuorenkoski A (2015) Marine animal classification using combined cnn and hand-designed image features. In: OCEANS 2015-MTS/IEEE Washington. IEEE, pp. 1–6
4. Chen B, Wang J, Zhao H, Zheng N, Principe JC (2015) Convergence of a fixed-point algorithm under maximum correntropy criterion. IEEE Signal Process Lett 22(10):1723–1727
5. Cormen TH (2009) Introduction to algorithms. MIT Press
6. Cronkite G, Enzenhofer HJ, Holmes JA, Fisheries D, Oceans Nanaimo BSB (2008) Evaluation of the blueview proviewer 900 imaging sonar as a tool for counting adult sockeye salmon in the Adams river, British Columbia. Tech. rep., DFO. Nanaimo
7. Dalgleish F (2013) Extended range distributed serial-scan laser imaging in turbid coastal conditions high-resolution laser imagery transmittable from seabed to surface. Sea Technol 54(12):15–+
8. Dalgleish F, Vuorenkoski A, Nootz G, Ouyang B, Caimi F (2013) Experimental study into the performance impact of the environmental noise on undersea pulsed laser serial imagers. J Underwater Acoust (USN) 61(4)
9. Dalgleish FR, Vuorenkoski AK, Ouyang B (2013) Extended-range undersea laser imaging: Current research status and a glimpse at future technologies. Mar Technol Soc J 47(5):128–147
10. Donahue J, Jia Y, Vinyals O, Hoffman J, Zhang N, Tzeng E, Darrell T (2014) Decaf: a deep convolutional activation feature for generic visual recognition. In: ICML, pp 647–655
11. Donoser M, Riemenschneider H, Bischof H (2009) Efficient partial shape matching of outer contours. In: Asian conference on computer vision. Springer, pp 281–292
12. Giddings TE, Shirron JJ, Tirat-Gefen A (2005) Eodes-3: an electro-optic imaging and performance prediction model. In: Proceedings of OCEANS 2005 MTS/IEEE. IEEE, pp 1380–1387
13. Hasanbelliu E, Giraldo LS, Principe JC (2014) Information theoretic shape matching. PAMI 36(12)
14. Joslin J, Rush B, Stewart A, Polagye B et al (2014) Development of an adaptable monitoring package for marine renewable energy projects part ii: hydrodynamic performance
15. Ling H, Jacobs DW (2007) Shape classification using the innerdistance. PAMI 29(2):286–299
16. Liu W, Pokharel PP, Principe JC (2007) Correntropy: properties and applications in non-gaussian signal processing. IEEE Trans Signal Process 55(11):5286–5298
17. Ma J, Qiu W, Zhao J, Ma Y, Yuille AL, Tu Z (2015) Robust l2e estimation of transformation for non-rigid registration. IEEE Trans Signal Process 63(5):1115–1129
18. Ma J, Zhao J, Yuille AL (2016) Non-rigid point set registration by preserving global and local structures. IEEE Trans Image Process 25(1):53–64
19. Magowan K, Reitsma J, Murphy D (2012) Use of dual-frequency identification sonar to monitor adult river herring in a small coastal stream. Marine Coastal Fish 4(1):651–659
20. Mobley CD (1994) Light and water: radiative transfer in natural waters. Academic Press
21. Ouyang B (2007) Watermarking based on unified pattern recognition framework. Ph.D. thesis, Southern Methodist University
22. Ouyang B, Dalgleish F, Caimi F, Vuorenkoski A, Giddings T, Shirron J (2012) Image enhancement for underwater pulsed laser line scan imaging system. In: SPIE Defense, security, and sensing. International Society for Optics and Photonics, pp 83,720r–83,720r
23. Ouyang B, Dalgleish F, Vuorenkoski A, Britton W, Ramos B, Metzger B (2013) Visualization and image enhancement for multistatic underwater laser line scan system using image-based rendering. IEEE J Ocean Eng 38(3):566–580
24. Polagye B, Copping A, Suryan R, Kramer S, Brown-Saracino J, Smith C (2014) Instrumentation for monitoring around marine renewable energy converters: workshop final report. PNNL-23110 Pacific Northwest National Laboratory. Seattle
25. Rother C, Kolmogorov V, Blake A (2004) Grabcut–interactive foreground extraction using iterated graph cuts. ACM Trans Graph 23(3):309–314
26. Sebastian TB, Klein PN, Kimia BB (2004) Recognition of shapes by editing their shock graphs. IEEE Trans Pattern Anal Mach Intell 26(5):550–571
27. Wang J, Bai X, You X, Liu W, Latecki LJ (2012) Shape matching and classification using height functions. Pattern Recogn Lett 33(2):134–143



Zheng Cao is currently a PhD student in the Computational NeuroEngineering Laboratory (CNEL) at the University of Florida. He received his BS degree from Nanjing University of Science and Technology in 2010, and MS degree from University of Wyoming in 2012. His current research interests are machine learning and computer vision, with applications in marine animal detection and classification. He has previously worked on system identification and kernel adaptive filtering.

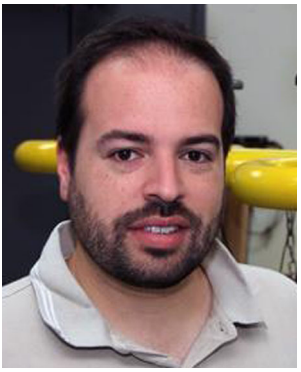
José C. Príncipe is a Distinguished Professor of Electrical and Computer Engineering and Biomedical Engineering at the University of Florida. He is also the BellSouth Professor and Founding Director of Computational NeuroEngineering Laboratory (CNEL), University of Florida. His primary research interests are advanced signal processing with information theoretic criteria (entropy and mutual information), adaptive models in the reproducing kernel Hilbert spaces (RKHS) and the application of these advanced algorithms in Brain Machine Interfaces (BMI). Dr. Principe is a Fellow of the IEEE, ABME and AIBME. He is the past Editor-in-Chief of the IEEE Transactions on Biomedical Engineering, past Chair of the Technical Committee on Neural Networks of the IEEE Signal Processing Society and past President of the International Neural Network Society. He received the IEEE EMBS Career Award, and the IEEE Neural Network Pioneer Award. He has more than 600 publications and 30 patents.

Bing Ouyang (S'02-M'06) received the Ph.D. degree in electrical engineering from Southern Methodist University, Dallas, TX, USA, in 2007. He joined the Ocean Visibility and Optics Laboratory, Harbor Branch Oceanographic Institute (HBOI), Florida Atlantic University (FAU), Fort Pierce, FL, USA, in 2009. Before joining HBOI, he was with Texas Instruments, Inc. (TI), Dallas, TX, USA. From 2003 to 2009, he was an Algorithm Engineer with the DLP ASIC algorithm team, with the primary focus in developing front end algorithms for the video processing ASIC. He holds four U.S. patents in the area of analog video and graphics format detection. His current research interests include underwater computer vision, novel underwater electro-optical system design, underwater LIDAR imaging enhancement, pattern recognition, and analysis for sensor time series data. Dr. Ouyang is a recipient of the 2013 Young Investigator Research Program award. He was peer elected to the Member of Technical Staff while working at TI.



Fraser Dalgleish received the Ph.D. degree in ocean engineering from Cranfield University, Cranfield, Bedford, U.K., in 2004. He directs the Ocean Visibility and Optics Laboratory, Harbor Branch Oceanographic Institute (HBOI), Florida Atlantic University (FAU), Fort Pierce, FL, USA. His research emphasis is on undersea optical sensor development, both for remote and in situ environmental measurements and to improve sensing and communications capabilities as an enabling technology for multivehicle imaging and sensing operations. Recent development activities in collaboration with government and industry have focused on new laser instrumentation, simulation tools, and detection approaches for ocean sensing and monitoring applications. Projects within academia involve collaboration with scientists in sensor package development and ocean observatories for water quality monitoring and imaging missions.

Anni Vuorenkoski received the Ph.D. degree in mechanical engineering from Cranfield University, Cranfield, Bedford, U.K., in 2004. She is with the Ocean Visibility and Optics Laboratory, Harbor Branch Oceanographic Institute (HBOI), Florida Atlantic University (FAU), Fort Pierce, FL, USA. Her current research concentrates on the characterization of water column and benthic features by polarimetry, fluorometry, angularly resolved scattering techniques, and time-resolved short-pulse methods, as well as on the experimental validation of computational radiative transfer models. Her past research activities have included adoption of optical, laser-based techniques to study the properties of turbid media and flows. She has also been involved in the development and application of Monte-Carlo-based computational methods to simulate the effects of multiple scattering in aerosol laser imaging.



Brian Ramos (M'12) received the B.S. degree in computer engineering from Florida Institute of Technology, Melbourne, FL, USA. He is currently working toward the M.S. degree in electrical engineering at Florida Atlantic University (FAU), Fort Pierce, FL, USA. He has worked for Harbor Branch Oceanographic Institute (HBOI), FAU, for over ten years, most recently focused on underwater imaging and communications.

Gabriel Alsenas received his B.S. and M.S. degree in Ocean and Systems Engineering from Florida Atlantic University (FAU), in 2005 and 2007 respectively. He has been the program manager at FAU Southeast National Marine Renewable Energy Center for 7 years.