Identifying sea scallops from benthic camera images

Prasanna Kannappan *

Justin H. Walker †

Art Trembanis ‡

Herbert G. Tanner §

*Department of Mechanical Engineering, University of Delaware Newark, DE 19716

†Department of Geological Sciences, University of Delaware, Newark, DE 19716

‡Department of Geological Sciences, University of Delaware, Newark, DE 19716

§Department of Mechanical Engineering, University of Delaware, Newark, DE 19716
Corresponding author. Email:btanner@udel.edu

Automated Scallop Counting from Images

Acknowledgments

The authors extend thanks to the captain and crew of the F/V CHRISTIAN AND ALEXA. This work grew out of a project with Dr. Bill Phoel (deceased) and his guiding vision and enthusiasm has carried this work forward. Special thanks are also extended to Scott Gallager and Amber York of Woods Hole Oceanographic Institution for making available to us datasets on which Matthew Dawkings and Charles Stewart of RPI's Computer Science Department worked on. The autonomous underwater vehicle field effort was funded by the NOAA Research Set-Aside Program under Award Number: NA11NMF4540011, and the automated image analysis work was supported by the National Science Foundation under IIS grant #0913015.

Abstract

The paper presents an algorithmic framework for the automated analysis of benthic imagery data collected by an autonomous underwater vehicle for the purpose of population assessment of epibenthic organisms, such as scallops. The architecture consists of three layers of processing. They are based on computational models of visual attention, graph-cut segmentation methods, and template matching, respectively. The visual attention layer filters the imagery input, focusing subsequent processing only on regions in the images that are likely to contain target objects. The segmentation layer prepares for subsequent template matching, which in turn sets the stage for classification of filtered objects into targets and distractors. The significance of the proposed approach is in its modular nature and its ability to process imagery datasets of low resolution, low brightness, and contrast.

Introduction

Background and Scope

The sea scallop *(Placopecten magellanicus)* fishery in the US EEZ (Exclusive Economic Zone) of the northwest Atlantic Ocean has been, and still is, one of the most valuable fisheries in the United States. Historically, the inshore sea scallop fishing grounds in the New York Bight, i.e., Montauk Point, New York to Cape May, New Jersey, have provided a substantial amount of scallops (Caddy 1975; Serchuk et al. 1979; Hart and Rago 2006; Naidu and Robert 2006; Fisheries of the United States 2012). These mid-Atlantic Bight "open access" grounds are especially important, not only for vessels fishing in the day boat category, which are usually smaller vessels with limited range opportunities, but also all the vessels that want to fish in near-shore "open access" areas to save fuel. These areas offer high fish densities, but are at times rapidly depleted due to overfishing (Rosenberg 2003).

Dredge-based surveys have been extensively used for Scallop population density assessment (National Marine Fisheries Service Northeast Fisheries Science Center (NEFSC) 2010). This involves dredging a part of the ocean floor, and manually counting the animals of interest found in the collected material. Besides being very invasive and disturbing to the creatures' habitat (Jenkins et al. 2001), these methods have severe accuracy limitations and can only generalize population numbers up to a certain extent. The goal of this paper is to demonstrate (a) the efficacy of non-invasive techniques of monitoring and assessing such populations through the use of an Autonomous Underwater Vehicle (auv) (Trembanis et al. 2011), and (b) the potential for automated methods of detection and enumeration of scallops.

To accomplish this goal, we developed a scallop counting system that collects seafloor imagery data using an auv and then analyzes it using a novel combination of machine vision methods. Our analysis workflow uses visual attention to mark possible scallop regions, and then implements segmentation and classification methodologies. The following sections will describe the constituent components in the context of literature.

Robotic Marine Surveys

Optical based surveys of benthic habitats, either from towed camera sleds or underwater robots, have introduced a huge leap forward in terms of data density for habitat studies. However, the abundance in seabed images is both a tremendous boon and also a challenge for researchers and managers with limited staff and time, struggling to process and analyze several hundreds of thousands to millions of images. So far, the development of new image acquisition strategies and platforms have far outstripped the development of image processing techniques. This mismatch provides the motivation behind our effort to automate the detection of images containing scallops.

One of the earliest video based surveys of scallops (Rosenkranz et al. 2008) notes that it took from 4 to 10 hours of tedious manual analysis in order to review and process one hour of collected seabed imagery. The report goes on to suggest that an automated computer technique for processing of the benthic images would be a great leap forward but at that time—and to the present—no such system has been available. There is anecdotal evidence of in-house development efforts by the HabCam group (Gallager et al. 2005) towards an automated system but as yet no such system has emerged to the community of researchers and managers. A recent manual count of our auv-based imagery dataset indicated that it took an hour to process 2080 images, whereas expanding the analysis to include all benthic macro-organisms reduced the rate down to 600 images/hr (Walker 2013). Another manual counting effort (Oremland et al. 2008) reports a processing time of 1 to 10 hours per person to process each image tow transect (exact image number per tow not listed). The same report indicates that the processing time was reduced considerably to 1–2 hours per tow by counting only every one-hundredth image, i.e. subsampling 1 % of the images.

Selective Processing

Visual attention is a neuro-physiologically inspired machine learning method (Koch and Ullman 1985). It attempts to mimic the human brain function in its ability to rapidly single out objects in imagery data that are different from their surroundings. It is based on the hypothesis that the human visual system first isolates points of interest from an image, and then sequentially processes these points based on the degree of interest associated with each point. The degree of interest associated with a pixel is called *salience*. Points with high salience values are processed first. The method therefore can be used to pinpoint regions in an image where the value of some pixel attributes may be an indicator to its uniqueness relative to the rest of the image.

According to the visual attention hypothesis (Koch and Ullman 1985), in the human visual system the input video feed is split into several feature streams. Locations in these feature streams which are very different from their neighborhoods correspond to peaks in the *center-surround* feature maps (explained later in detail). The different center-surround feature maps can be combined to obtain a saliency *map*. Peaks in the saliency maps, otherwise known as *fixations*, are points of interest, processed sequentially in descending order of their salience values.

Itti et al. (1998) proposed a computational model for visual attention. According to this model, an image is first processed along three feature streams (color, intensity, and orientation). The color stream is further divided into two sub-streams (red-green and blue-yellow) and the orientation stream into four sub-streams ($\theta \in \{0°, 45°, 90°, 135°\}$). The image information in each sub-stream is further processes in 9 different scales. In each scale, the image is scaled down using a factor $\frac{1}{2^k}$ (where $k = 0, \ldots, 8$), resulting in some loss of information as scale increases. The resulting image data for each scale factor constitutes the *spatial scale* for the particular sub-stream.

The sub-stream feature maps are compared across different scales to expose differences in them. Though the spatial scales in each sub-stream feature map originated from the same map, the scaling factors change the information contained in each map. When these spatial scales are resized to a common scale through interpolation and compared to get center-surround feature maps, the mismatches between the scales get highlighted. For the intensity stream, the center-surround feature map is given by

$$I(c, s) = |I(c) \ominus I(s)| \quad , \tag{1}$$

where $\ominus$ is the *center-surround* operator that takes pixel-wise differences between resized sub-streams to exposes those mismatches, c and s are indices for two different spatial scales with $c \in \{2, 3, 4\}$, $s = c + \delta$, for $\delta \in \{3, 4\}$. Similarly center-surround feature maps are computed for each sub-stream in color and orientation streams.

The seven sub-streams (two in color, one in intensity and four in orientation), yield 42 center-surround feature maps. The center-surround feature maps in each original stream (color, intensity, and orientation) are then combined into three *conspicuity maps*: one for color $\bar{C}$, one for intensity $\bar{I}$, and one for orientation $\bar{O}$. For instance, the intensity conspicuity map is computed as below.

$$\bar{I} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c=4} w_{cs} \mathcal{N}(I(c,s)) \tag{2}$$

where the $\oplus$ cross-scale operator works in a fashion similar to $\ominus$, but the difference being that data in the resized maps from different scales is pixel-wise added. The map normalization operator $\mathcal{N}(\cdot)$ in (2) scales a map by the scaling factor $(M - \bar{m})^2$, where M is the global maximum over the map and $\bar{m}$ is the mean over all local maxima present in the map. Finally, the 3 conspicuity maps are combined to get a *saliency map*

$$S = w_{\bar{I}} \mathcal{N}(\bar{I}) + w_{\bar{C}} \mathcal{N}(\bar{C}) + w_{\bar{O}} \mathcal{N}(\bar{O}) \ , \tag{3}$$

where $w_{\bar{k}}$ is a user-selected stream weight. In Bottom-Up Visual Attention (buva) all streams are weighted equally, so $w_{\bar{I}} = w_{\bar{C}} = w_{\bar{O}} = 1$. On this saliency map, a winner-takes-all neural network is typically used (Itti et al. 1998; Walther and Koch 2006) to compute the maxima (without loss of generality, any other methods to compute maxima can be used). Visual attention methods call these local maxima as fixations, which lead to shifts in the focus of attention to these points. Visual attention explains focus of attention as sub-sampled regions in the image which the brain processes preferentially at some instant of time.

The weights in (2) and (3) can be selected judiciously to bias fixations toward specific targets of interest. The resulting variant of this method is known as Top-Down Visual

Attention (tdva) (Navalpakkam and Itti 2006). One method to select these weights is (Navalpakkam and Itti 2006):

$$w_j = \frac{w'_j}{\frac{1}{N_m}\sum_{j=1}^{N_m} w'_j} \quad , \tag{4a}$$

where N_m is the number of feature (or conspicuity) maps, and

$$w'_j = \frac{\sum_{i=1}^{N} N_{iT}^{-1} \sum_{k=1}^{N_{iT}} P_{ijT_k}}{\sum_{i=1}^{N} N_{iD}^{-1} \sum_{k=1}^{N_{iD}} P_{ijD_k}} \quad , \tag{4b}$$

where N is the number of images in the learning set, N_{rT} and N_{rD} are the number of targets (scallops) and distractors (similar objects) in the r-th learning image, P_{uvT_z} is the mean salience value of the region around the v-th map containing the z-th target (T) in the u-th image. P_{uvD_z} is similarly defined for distractors (D).

Vision-based Detection of Marine Creatures

There have been attempts to count marine species using stationary underwater cameras (Edgington et al. 2006; Spampinato et al. 2008). In this general framework, salmon are counted through background subtraction and shape detection (Williams et al. 2006). However, counting sedentary and sea-floor inhabiting animals like scallops does not come under the purview of these methods, since background subtraction is inherently challenging. In some setups, like those used for zooplankton assessment (Stelzer 2009; McGavigan 2012) very specialized imaging and sampling apparatus is required, which cannot be easily retasked for other applications. auvs with mounted cameras have been used for identification of creatures like clam and algae (Forrest et al. 2012). In such cases, very simple processing techniques like thresholding and color filtering are used. These techniques have little chance of success with scallops, as scallops do not exhibit any unique color or texture.

One way to approach the problem of detecting marine animals from seabed images is by detecting points of interest in an image, which are most likely to contain objects that differ significantly from their background. Singling out these regions of interest does not automatically produce positive counts, because a wealth of other features can trigger false positives. Additional processing of the region around the candidate points is needed to identify targets of interest. However, the detection method can be biased toward the features of the target, and thus reduce the number of false positives.

Technically, points of interest are locations in the datastream where there is a sudden change in the underlying distribution from which the data is generated. Some mathematical approaches to determining this change in distribution can be found in (Basseville and Nikiforov 1993; Poor and Hadjiliadis 2009). However most of these methods require some prior knowledge about the underlying distribution. Modeling the background distribution from image data can be problematic without several simplifying technical assumptions, sometimes of debatable validity in the specific application context.

Scallops, especially when viewed in low resolution, do not provide features that would clearly distinguish them from their natural environment. This presents a major challenge in automating the identification process based on visual data. To compound this problem, visual data collected from the species' natural habitat contain a significant amount of speckle noise. Some scallops are also partially or almost completely covered by sediment, obscuring the scallop shell features. A highly robust detection mechanism is required to overcome these impediments.

The existing approaches to automated scallop counting in artificial environments (Enomoto et al. 2009, 2010) employ a detection mechanism based on intricate distinguishing features like fluted patterns in scallop shells and exposed shell rim of scallops respectively. Imaging these intricate scallop shell features might be possible in artificial scallop beds with stationary cameras and minimal sensor noise, but this level of detail is difficult to obtain from images of scallops in their natural environment. A major factor that contributes to this loss in detail is the poor image resolution obtained when the image of the target is captured several meters away from it. Overcoming this problem by operating an underwater vehicle too close to the ocean floor will adversely impact the image footprint (i.e. area covered by an image) and also the drivability of the vehicle due to relief structures on the ocean floor.

The existing work on scallop detection (Dawkins 2011; Einar Óli Guòmundsson 2012) in their natural environment is limited to small datasets. From these studies alone, it is not clear if such methods can be used effectively in cases of large data sets comprising several thousand seabed images, collected from auv missions as the test sets used here are often less than 100 images. An interesting example of machine-learning methods applied to the problem of scallop detection (Fearn et al. 2007) utilizes the concept of buva. The approach is promising but it does not use any ground truth for validation. As with several machine learning and image processing algorithms, porting the method from the original application set-up to another may not necessarily yield the anticipated results, and the process has to be tested and assessed.

Contributions

The paper describes a combination of robotic-imaging marine survey methods, with automated image processing and detection algorithms. The automated scallop detection algorithm workflow involves 3 processing layers based on customized tdva pre-processing, robust image segmentation and object recognition methods respectively. The paper does not claim major innovations in the computational approach's constituent technologies; however, some degree of customization, fine-tuning and local improvement is introduced. The value of the proposed approach is primarily in the field application front, providing a novel engineering solution to a real-world problem with economic and societal significance, that goes beyond the particular domain of scallop population assessment and can possibly extend to other problems of environmental monitoring, or even defense (e.g. mine detection). Given the general unavailability of similar automation tools, the proposed one can have potential impact in the area of underwater automation. The multi-layered approach not only introduces several innovations at the implementation level, but also provides a specialized package for benthic habitat assessment. At a processing level it provides the flexibility to re-task individual data processing layers for different detection applications. When viewed as a complete package, the proposed approach offers an efficient alternative to benthic habitat specialists for processing large image datasets.

Materials and Procedure

The 2011 RSA project (Titled: "A Demonstration Sea Scallop Survey of the Federal Inshore the New York Bight using a Camera Mounted Autonomous Underwater Vehicle.") was a proof-of-concept project that successfully used a digital, rapid-fire camera integrated to a Gavia auv, to collect a continuous record of photographs for mosaicking, and subsequent scallop enumeration. In July 2011, transects were completed in the northwestern waters of the mid-Atlantic Bight at depths of 25-50 m. The auv continuously photographed the seafloor along each transect at a constant altitude of 2 m above the seafloor. Parallel sets of transects were spaced as close as 4 m, offering unprecedented two-dimensional spatial resolution of sea scallops. Georeferenced images were manually analyzed for the presence of sea scallops using position data logged (using Doppler Velocity Log (dvl) and Inertial Navigation System (ins)) with each image.

Field Survey Process

In the 2011 demonstration survey, the federal inshore scallop grounds from Shinnecock, New York to Ocean View, Delaware, was divided into eight blocks or strata (as shown in

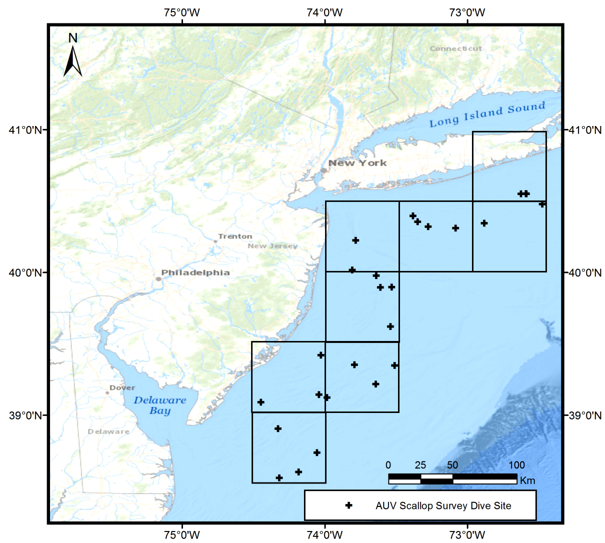


Figure 1: Map of the survey region from Shinnecock, New York to Cape May, New Jersey, divided into eight blocks or strata

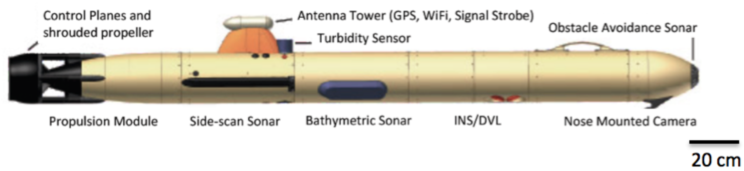
Figure 1). The F/V Christian and Alexa served as the surface support platform from which a Gavia AUV (see Figure 2) was deployed and recovered. The AUV conducted photographic surveys of the seabed for a continuous duration of approximately 3 hours during each dive, repeated 3–4 times in each stratum, with each stratum involving roughly 10 hours of imaging and an area of about $45\,000\ \mathtt{m}^2$. The AUV collected altitude (height above the seabed) and attitude (heading, pitch, roll) data, allowing the georectification of each image into scaled images for size and counting measurements. During the 2011 pilot study survey season, over 250 000 images of the seabed were collected. These images were analyzed in the University of Delaware's laboratory for estimates of abundance and size distribution. The F/V Christian and Alexa provided surface support, and made tows along the AUV transect to ground-truth the presence of scallops and provide calibration for the size distribution. Abundance and sizing estimates were conducted via a heads-up manual method, with each image including embedded metadata allowing it to be incorporated into to existing benthic image classification systems (HabCam MIP (Dawkins et al. 2013)).

During this proof of concept study, in each stratum the f/v Christian and Alexa made one 15-minute dredge tow along the auv transect to ground-truth the presence of scallops and other fauna, and provide calibration for the size distribution. The vessel was maintained on the dredge track by using Differential dGlobal Positioning System (gps). The tows were made with the starboard 15 ft (4.572 m) wide New Bedford style commercial dredge at the commercial dredge speed of 4.5–5.0 knots. The dredge was equipped with 4 inch (10.16 m) interlocking rings, an 11 inch (27.94 cm) twine mesh top, and turtle chains. After dredging, the catch was sorted, identified, and weighed. Length-frequency data were obtained for the caught scallops. This information was recorded onto data logs and then entered into a laptop computer database aboard ship for comparison to the camera image estimates.

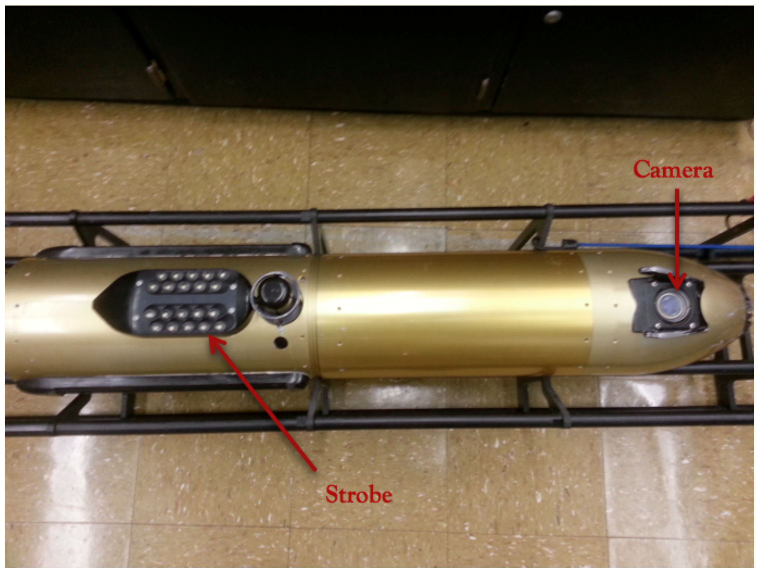
The mobile platform of the auv provided a more expansive and continuous coverage of the seabed compared to traditional fixed drop camera systems or towed camera systems. In a given day, the auv surveys covered about 60 000 m^2 of seabed from an altitude of 2 m above the bed, simultaneously producing broad sonar swath coverage and measuring the salinity, temperature, dissolved oxygen, and chlorophyll-a in the water.

Sensors and Hardware

The University of Delaware auv (Figure 2) was used to collect continuous images of the benthos, and simultaneously map the texture and topography of the seabed. Sensor systems associated with this vehicle include: (1) a 500 kHz GeoAcoustics GeoSwath Plus phase measuring bathymetric sonar; (2) a 900/1800 kHz Marine Sonic dual-frequency high-resolution side-scan sonar; (3) a Teledyne RD Instruments 1200 kHz acoustic doppler velocity log (DVL)/Acoustic doppler current profiler (ADCP); (4) a Kearfott T-24 inertial navigation system; (5) an Ecopuck FLNTU combination fluorometer / turbidity sensor; (6) a



(a)



(b)

Figure 2: Schematics and image of Gavia auv

Point Grey Scorpion model 20SO digital camera and LED strobe array; (7) an Aanderaa Optode dissolved oxygen sensor; (8) a temperature and density sensor; and, (9) an altimeter. Each sensor separately records time and spatially stamped data with frequency and spacing. The AUV is capable of very precise dynamic positioning, adjusting to the variable topography of the seabed while maintaining a constant commanded altitude offset.

Data Collection

The data was collected over two separate five-day cruises in July 2011. In total, 27 missions were run using the auv to photograph the seafloor (For list of missions see Table 1). Mission lengths were constrained by the 2.5 to 3.5 hour battery life of the auv. During each mission, the auv was instructed to follow a constant height of 2 m above the seafloor. In addition to the 250 000 images that were collected, the auv also gathered data about water temperature, salinity, dissolved oxygen, geoswath bathymetry, and side-scan sonar of the seafloor.

The camera on the auv, a Point Grey Scorpion model 20SO (for camera specifications see Table 2), was mounted inside the nose module of the vehicle. It was focused at 2 m, and captured images at a resolution of 800×600. The camera lens had a horizontal viewing angle of 44.65 degrees. Given the viewing angle and distance from the seafloor, the image footprint can be calculated as $1.86 \times 1.40\ \mathtt{m}^2$. Each image was saved in JPEG format, with metadata that included position information (including latitude, longitude, depth, altitude, pitch, heading and roll) and the near-seafloor environmental conditions analyzed in this study. This information is stored in the header file, making the images readily comparable and able to be incorporated into existing RSA image databases, such as the HabCam database. A *manual* count of the number of scallops in each image was performed and used to obtain overall scallop abundance assessment. Scallops counted were articulated shells in life position (left valve up) (Walker 2013).

Layer I: Top-Down Visual Attention

Counting the scallops manually through observation and tagging of the auv-based imagery dataset, is a tedious process that typically proceeds at a rate of 600 images/hr (Walker 2013). The outcome usually includes an error in the order of 5 to 10 percent. An automated system that would just match this performance would still be preferable to the arduous manual process.

Classification methods generally depend on some characteristic features of objects of interest. The selection of features on scallops is an issue that can be open to debate, and different suggestions can be given depending on context. Our dataset, (see Figure 3 for a

Table 1: List of missions and number of images collected

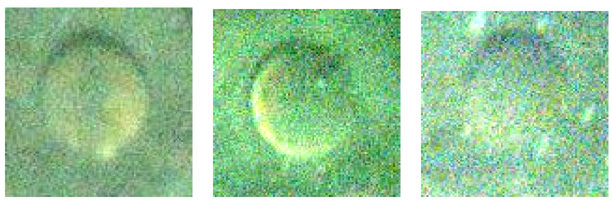
Mission	Number of images	Mission	Number of images
LI1[1]	12 775	NYB6	9 281
LI2	2 387	NYB7	12 068
LI3	8 065	NYB8	9 527
LI4	9 992	NYB9	10 950
LI5	8 338	NYB10	9 170
LI6	11 329	NYB11	10 391
LI7	10 163	NYB12	7 345
LI8	9 780	NYB13	6 285
LI9	2 686	NYB14	9 437
NYB1[2]	9 141	NYB15	11 097
NYB2	9 523	ET1[3]	9 255
NYB3	9 544	ET2	12 035
NYB4	9 074	ET3	10 474
NYB5	9 425		

[1] LI–Long Island
[2] NYB–New York Bight
[3] ET–Elephant Trunk

Table 2: Camera specifications

Attribute	Specs
Name	Point Grey Scorpion 20SO Low Light Research Camera
Image Sensor	8.923 mm Sony CCD
Horizontal Viewing Angle	44.65 degrees (underwater)
Mass	125 g
Frame rate	3.75 fps
Memory	Computer housed in AUV nose cone
Image Resolution	800 × 600
Georeferenced metadata	Latitude, longitude, altitude, depth
Image Format	JPEG

Figure 3: Seabed image with scallops shown in circles



(a) (b) (c)

Figure 4: (a) Scallop with yellowish tinge and dark crescent; (b) Scallop with yellowish tinge and bright shell rim crescent; (c) Scallop with no prominent crescents and texturally identical to the background

representative sample) does not offer any unequivocal feature choices, but there were some identifiable recurring visual patterns.

One example is a dark crescent on the upper perimeter of the scallop shell, which is the shadow cast by the upper open scallop shell produced from the auv strobe light (see Figure 4(a)). Another pattern that could serve as a scallop feature in this dataset is a frequently occurring bright crescent on the periphery of the scallop, generally being the visible inside of the right (bottom) valve when the scallop shell is partly open (see Figure 4(b)). A third pattern is a yellowish tinge associated with the composition of the scallop image (see Figure 4(b)).

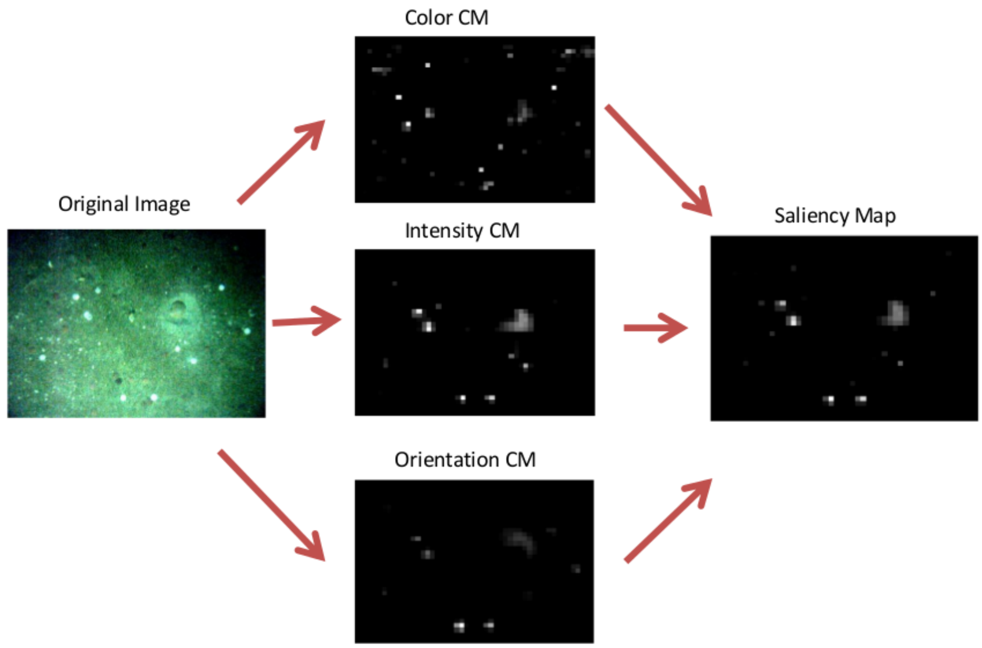


Figure 5: Illustration of saliency map computation

Learning

A tdva algorithm was customized to sift automatically through the large volume of imagery data, and focus on regions of interest that are more likely to contain scallops. First, bottom-up saliency computation is performed on 243 annotated images, collectively containing 300 scallops (see Figure 5). Figure 5 illustrates the process of computing the color, intensity, and orientation conspicuity maps from the original image. These conspicuity maps are subsequently combined to yield the saliency map. The intermediate step of computing the center-surround feature maps has been omitted from the figure for the sake of clarity. In each saliency map, fixations are identified through a process of extremum seeking that identifies the highest saliency values. In Figure 6, the yellow outline around the annotated peaks is the *proto-object* (Walther and Koch 2006). From empirical observation, these proto-objects rarely contain scallops; they are usually regions texturally identical to the fixation point. The fixation points often occur near the scallop boundary, but outside the

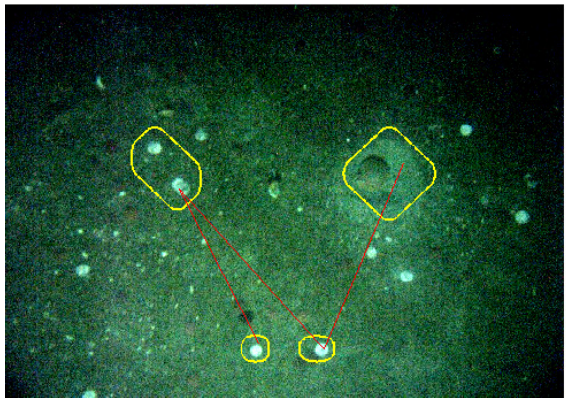


Figure 6: Illustration of fixations. The red lines indicate the order in which the fixations were detected with the lower-left fixation being the first. The yellow outline is the proto-object around the fixation.

scallop. This can be justified by the fact that typically in our images the center of the scallop is texturally identical to the background. Throughout this learning phase, the fixation window used is a rectangular window of size 100×100 pixels (approximately 23×23 $\texttt{cm}^2$ of seafloor) centered around fixation points. If the center of a scallop lies within this window, the corresponding fixation is labeled a *target*, and a *distractor* otherwise.

The target and distractor regions were determined in all the feature and conspicuity maps for each one of these processed images in the learning set. This is done by adaptively thresholding and locally segmenting the points around the fixations with similar salience values in each map. Then the mean of the salience values of these target and distractor regions from the feature maps and conspicuity maps is used to compute the top-down weights for feature maps and conspicuity maps, respectively, using (4).

The resulting top-down conspicuity map weights are $w_{\bar{I}} = 1.1644$, $w_{\bar{C}} = 1.4354$ and $w_{\bar{O}} = 0.4001$. The small value of the orientation weight is understandable, because scallops are for the most part symmetric and circular (This may not be true for high resolution photographs of scallop shells where the auricles and hinge would be much more prominent, but true for the low resolution dataset obtained from our survey.) The set of feature map

Table 3: Top-down weights for feature maps

		Center Surround Feature Scales					
		1	2	3	4	5	6
Color	red-green	0.8191	0.8031	0.9184	0.8213	0.8696	0.7076
	blue-yellow	1.1312	1.1369	1.3266	1.2030	1.2833	0.9799
Intensity	intensity	0.7485	0.8009	0.9063	1.0765	1.3111	1.1567
Orientation	0°	0.7408	0.2448	0.2410	0.2788	0.3767	2.6826
	45°	0.7379	0.4046	0.4767	0.3910	0.7125	2.2325
	90°	0.6184	0.5957	0.5406	1.2027	2.0312	2.1879
	135°	0.8041	0.6036	0.7420	1.5624	1.1956	2.3958

weights for each center-surround scale w_{cs} for every feature is listed in Table 3.

Testing and Implementation

During the testing phase, saliency maps are computed for images in the two datasets shown in Table 4. The saliency map computation involves using the top-down conspicuity weights given above and the feature map weights of Table 3 in (3) and (2).

Dynamic thresholds are employed to compute fixations from the saliency maps in this version of tdva. This mechanism controls the convergence time required for the winner-takes-all neural network, implemented for detecting fixations, i.e. peaks in the saliency map. It is highly unlikely that a fixation that contains an object of interest requires a convergence time of more than 10 000 iterations. In principle, even specks of noise can produce fixations if this neural network is allowed to evolve indefinitely. Dynamic threshold ensures that if convergence to some fixation takes more than this number of iterations, then the search is terminated and no more fixations are sought in the image.

At most ten fixations in each image are recorded in the decreasing order of their salience values. Ten fixations is deemed sufficient, given that there is an average of roughly

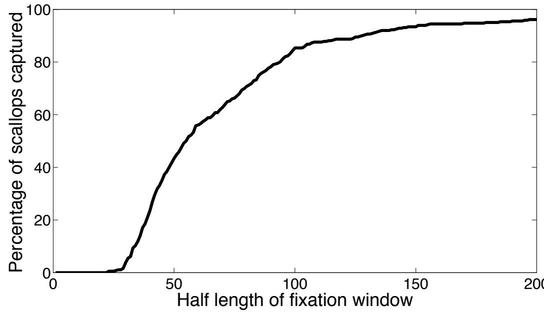


Figure 7: Percentage of scallops enclosed in the fixation window as a function of window half length (in pixels)

(a) (b) (c) (d) (e)

Figure 8: (a) Fixation window from layer I; (b) Edge segmented image; (c) graph-cut segmented image; (d) Region boundaries obtained when the edge segmented image is used as a mask over the graph-cut segmented image boundaries; (e) circle fitted on the extracted region boundaries.

two scallops per image, and very few images contain more than ten scallops (5 images contained more than 10 scallops; that was 0.002% of the total images). The fixation window size in testing phase is enlarged to 270×270 pixels (approximately 63×63 $\mathtt{cm}^2$)—half window length of 135 pixels, because in testing phase the fixation window should be large enough to enclose the complete scallop and not just the scallop center, as required before in the learning phase. The chosen window size can enclose more than 91% of the scallops in the images, which have radii that vary between 20 and 70 pixels in our dataset (see Figure 7).

Layer II: Segmentation and Detection Criteria

Layer II comprises image segmentation algorithms that operate on the fixation windows obtained as a result of Layer I processing. This layer consists of three separate sub-layers: edge based segmentation (involves basic morphological operations like smoothing, adaptive thresholding and edge detection), graph-cut segmentation, and shape extraction. The segmentation process flow for a sample fixation window featuring a scallop is illustrated in Figure 8. Edge based segmentation on the fixation window of Figure 8(a) yields the edge segmented image of Figure 8(b). Figure 8 shows the effect of edge based segmentation and graph-cut segmentation on a fixation window, and also shows the shape fitting applied to the boundary contours obtained by combining edge based segmentation and graph-cut segmentation results.

The graph-cut segmentation sublayer extracts ten regions in each fixation window, transforming the window of Figure 8(a) to the segmented image shown in Figure 8(c). In this approach, the image segmentation problem is reduced into a graph partition problem (Shi and Malik 2000). The graph $G = (V, E)$, with node set V and edge set E, consists of nodes associated with image pixels and edges being links between these nodes. Each edge $(u, v) \in E$ is assigned a weight $w(u, v)$, to form the weighted graph G. The weights on edges are assigned based on image features, and are computed as follows.

$$w(u,v) = \begin{cases} \exp\left(-\frac{\|F(u)-F(v)\|_2^2}{\sigma_I} - \frac{\|X(u)-X(v)\|_2^2}{\sigma_X}\right) \ , & \text{if } \|X(u) - X(v)\|_2 < r. \\ 0 \ , & \text{otherwise} \end{cases}$$

where $X(u)$ is the spatial coordinates of node u, $F(u)$ is the feature value vector (e.g. intensity, color, texture) at node u, r is a small positive threshold constant, and σ_I, σ_X are positive constants, selected typically within 10–20% of the range of feature values and spatial distances, respectively. Function $\| \cdot \|_2$ is the Euclidean norm.

The graph's nodes are partitioned into background nodes A, and foreground nodes B. This partitioning is obtained by solving an optimization problem that minimizes a *normalized* graph-cut function shown in (5). In other words, the partitioning works through the selection of (minimal) weights w on edges that link nodes background and foreground partitions. The methodology followed here is discussed in detail in Shi and Malik (2000).

$$\mathsf{Ncut}(A,B) = \frac{\sum_{u\in A, v\in B} w(u,v)}{\sum_{p\in A, q\in V} w(p,q)} + \frac{\sum_{u\in A, v\in B} w(u,v)}{\sum_{p\in B, q\in V} w(p,q)} \ . \tag{5}$$

The partitioning process can be applied to cases where k partition blocks, $A_1, \ldots, A_k$, are required, by extending (6) to the objective function

$$\mathsf{Ncut}_k(A,B) = \frac{\sum_{u\in A_1, v\in V-A_1} w(u,v)}{\sum_{p\in A_1, q\in V} w(p,q)} + \cdots + \frac{\sum_{u\in A_k, v\in V-A_k} w(u,v)}{\sum_{p\in A_k, q\in V} w(p,q)} \ . \tag{6}$$

The shape extraction sublayer involves the fitting of a circle to a connected contour produced by the graph-cut segmentation sublayer (Figure 8(e)). The choice of the shape to be fitted is suggested by the geometry of the scallop's shell. Finding the circle that fits best to a given set of points can be formulated as an optimization problem (Taubin 1991; Chernov 2010). Given a set of n points with coordinates (x_i, y_i) with $i = 1, 2, \ldots, n$, an objective function to be minimized can be defined with respect to three design parameters, (a, b) and R—the center coordinates and the radius of the circle to be fitted—in the form

$$F_1(a,b,R) = \sum_{i=1}^{n} \left[(x_i - a)^2 + (y_i - b)^2 - R^2\right]^2 \ . \tag{7}$$

With this being the basic idea, it is shown (Taubin 1991) that a variation of (7) in the form

$$F_2(A,B,C,D) = \frac{\sum_{i=1}^{n} (Az_i + Bx_i + Cy_i + D)^2}{n^{-1}\sum_{i=1}^{n} (4A^2 z_i + 4ABx_i + 4ACy_i + B^2 + C^2)} \tag{8}$$

with the following re-parameterization

$$a = -\frac{B}{2A} \ , \qquad b = -\frac{C}{2A} \ , \qquad R = \sqrt{\frac{B^2 + C^2 - 4AD}{4A^2}} \ , \qquad z_i = x_i^2 + y_i^2 \ ,$$

yields the same solution for (a, b, R).

Once the circle is fit on the contour, the quality of the fit and its acceptance with the manually annotated scallop measurements is quantified. For this quantification, two measures that capture the error in center e_c and error percent in radius e_r of the fitted circle to that of the manually annotated scallop are defined in (9).

$$e_c = \sqrt{(a_g - a_s)^2 + (b_g - b_s)^2} \leq 12 \ \ (pixels) \tag{9a}$$

$$e_r = \frac{|R_g - R_s|}{R_g} \leq 0.3 \ \ . \tag{9b}$$

where the annotated scallop is represented by the triple (a_g, b_g, R_g)– coordinates of the center (a_g and b_g) and the radius R_g. Similarly (a_s, b_s, R_s) refers to the fitted circle. All measurements here are in pixels.

If both the errors are within specified thresholds, the scallop is considered to be successfully detected. The specific thresholds shown in (9) were set empirically, taking into account that the radius measurements in manual counts in (Walker 2013) (used as ground truth here) have a measurement error of 5–10 %.

Layer III: Classification

Layer III classifies the fitted circles from Layer II into scallops and non-scallops. This binary classification problem depends on identifying some specific markers that are unique to scallops. One such characteristic that was empirically identified from the images of scallops is the presence of two visible crescents, a bright crescent toward the lower periphery and a dark crescent toward the upper periphery. It is observed that these crescents appear on diametrically opposite sides. Though these are not part of the organism itself, but rather an artifact of the sensing system, they still provide specific information that can be exploited by the classification algorithm.

The sensing mechanism in the experimental setup contains a camera at the nose of the AUV, and a strobe light close to its tail (mounted to the hull of the control module at an oblique angle to the camera). Objects that rise above the seafloor exhibit a bright region closest to the strobe light and a dark shadow farthest away from the strobe light. These light artifacts combined with characteristic shape of scallop shell produce the visible crescents which were used to identify scallops.

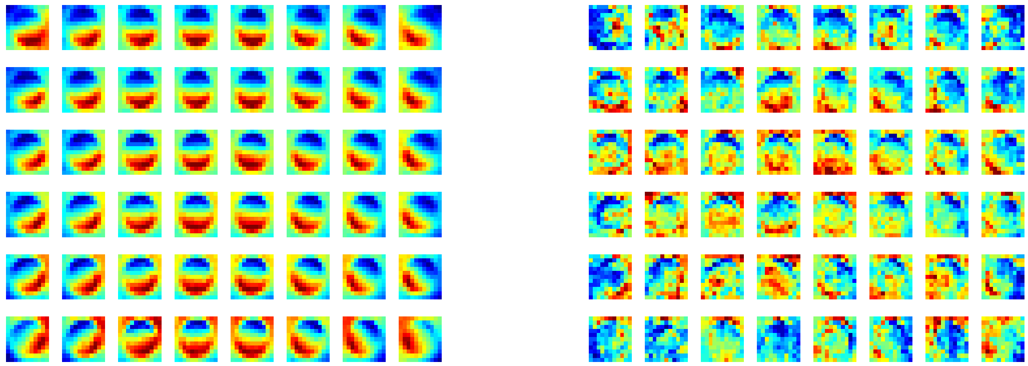
Though crescents appear in images of most scallops, their prominence and relative position with respect to the scallop varies considerably. Our hypothesis with regards to the origin of these light artifacts suggests that their properties are a function of the center pixel location on the image. If our hypothesis is true, the approximate profile of a scallop located at any point in the image can be pre-computed. These pre-computed profiles can then be compared with the objects obtained from the segmentation layer (Layer II). The shape, size, and orientation of these crescents can thus be indicative of the presence of a scallop at these locations, and such an indication can be quantified numerically using template matching methods.

Scallop Profile Hypothesis

To validate the hypothesis that the image profile of a scallop (shape and orientation of crescents) is dependent on its spatial location in the image, a statistical analysis was performed on a dataset of 3706 manually labeled scallops (each scallop is represented as (a, b, R) where a, b are the horizontal and vertical coordinates of the scallop center, and R is its radius). For this analysis, square windows of length $2.8 \times R$ centered on (a, b) were used to crop out regions from the images containing scallops. Using a slightly larger window size (size greater than $2 \times R$, the size of the scallop) includes the neighborhood pixels just outside the scallop pixels into the crop window (this was done to include the scallop crescent pixels which often appeared outside the scallop circle). The cropped scallop regions were then reduced to grayscale images and enhanced through contrast stretching, followed by binning the scallop pixels based on their spatial location. The slightly larger diameter ($2.8 \times R$ instead of $2 \times R$) also improves the performance of local contrast stretching which in turn strengthens the scallop boundaries. Since cropped scallops images can be of different sizes, they are normalized by resizing each of them to an 11×11 dimension. To demonstrate the dependence of the scallop profile on the pixel coordinates of its center point, the 600×800 image area(original image size of the dataset) is discretized into 48 bins (8 in horizontal direction, 6 in vertical direction, bin size 100×100). The scallops with centers that fall in each of these bins are segregated. Technically, each of the resulting 11×11 pixel images of scallops can be represented as a 121 dimensional vector. The mean and standard deviation maps of the scallop points in each bin are shown in Figure 9. The mean maps in Figure 9(a), illustrate the dependence of the scallop crescents on its position in the image. Additionally, the standard deviation maps in Figure 9(b) show that the *darker* crescent towards the top of the scallop is more consistent as a marker than the bright crescent, due to the relatively lower standard deviation of the former.

Scallop Profile Learning

The visual scallop signatures as a function of its spatial location on the image plane can be captured in form of a look-up table to streamline the classification process. The lookup table is constructed using the same dataset of 3 706 manually labeled scallops, that was used for the scallop profile hypothesis validation. For each pixel location in the 600×800 image, a mean and a standard deviation map (similar to the ones in Figure 9) is computed from

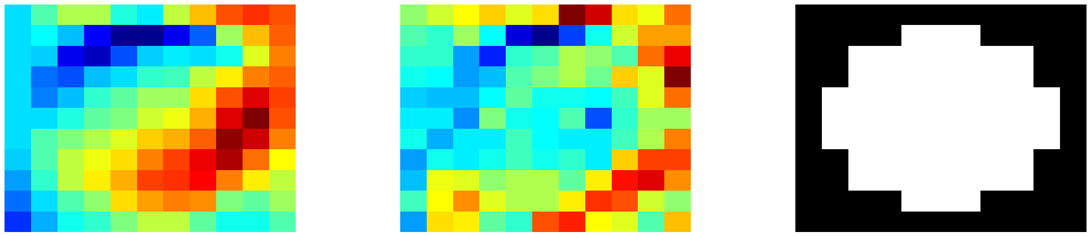


(a) (b)

Figure 9: (a) Mean map of scallops in each quadrant (b) Standard deviation map of scallops in each quadrant. Red corresponds to higher numeric values and blue correspond to lower numeric values.

scallops with centers lying within a 40×40 window centered on the pixel. After normalization (as done in scallop profile hypothesis verification procedure), the mean map results in a 121 dimensional feature vector (11×11) corresponding to each point in the the 600×800 image. Similar processing is then done for standard deviation maps. Both mean and standard deviation maps are stored onto the lookup table.

Although the feature vectors of Figure 9(a) may appear visually informative, not all 121 features are useful. This is because the maps for the mean were created using a radius around each pixel that is larger than the scallop radius. The implication of this is that the pixels close to the boundary of the 11×11 window containing the mean and standard deviation maps correspond to points that express background and thus do not contain relevant information. Thus a circular mask is applied to the maps, where the mask is centered on the 11×11 map and is of radius 4 pixels (equal to the average scallop radius). Figure 10 shows an instance of the data stored in the lookup table for a specific point with pixel coordinates $(\text{row}, \text{column}) = (470, 63)$ along with the circular mask. Application of this mask effectively reduces the number of features to 61. Considering that the standard



(a) (b) (c)

Figure 10: (a) Mean map of scallop; (b) Standard deviation map of scallop at point (row,column)=(470,63); (c) Mask applied to remove background points.

deviation of a feature is inversely proportional to its relative strength or importance, an additional 25% of the remaining features (15 features) having the highest standard deviation is ignored. These ignored features typically point to outliers and hinder subsequent template matching. With this, the number of features used to describe an identified object drops to 46.

Scallop Template Matching

Each object passed down from the segmentation layer (Layer II) is first cropped, and basic image processing steps as discussed in the scallop profile extraction process (in Layer III) are applied to obtain an 11×11 cropped object image. To be consistent with the scallop profile learning procedure, a crop window size of $2.8 \times R$ is used for cropping objects. The resulting 46-dimensional object feature vector is used for comparison with the reference scallop feature vector for template matching.

The 46-dimensional object point is normalized and then a comparison metric is computed. This comparison metric is a weighted distance function between the object point and the reference scallop profile at that point. If this distance metric is greater than a certain threshold, the object is not counted as a scallop, otherwise it is considered a scallop. Technically, if $X^o = (X_1^o, X_2^o, \ldots, X_{46}^o)$ denotes the object point and $X^s = (X_1^s, \ldots, X_{46}^s)$ the corresponding reference scallop profile, then the component at location p in the normalized

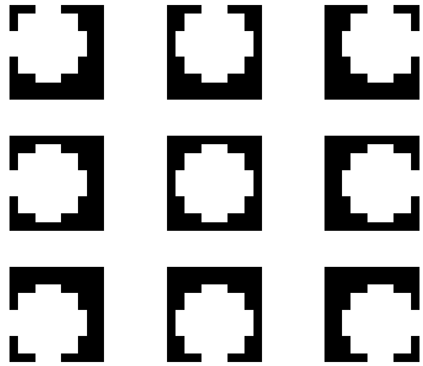


Figure 11: Nine different masks slightly offset from the center used to make the classification layer robust to errors in segmentation

object feature vector $X^{\bar{o}}$ is given by

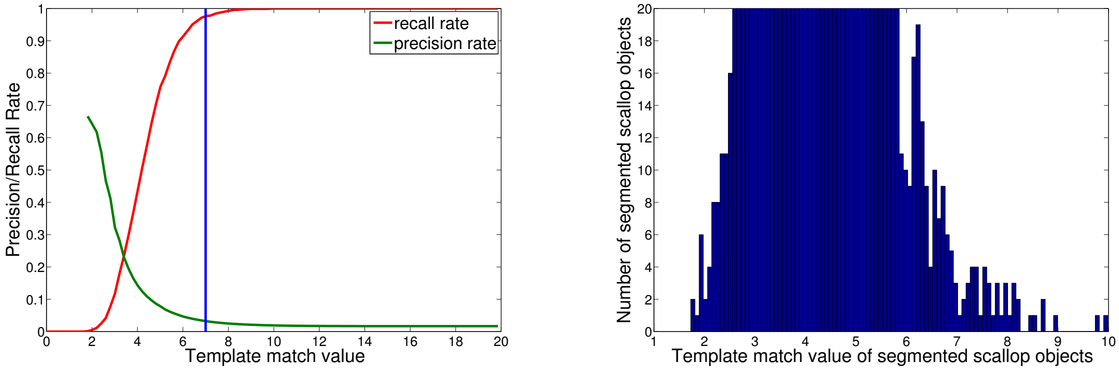
$$X_p^{\bar{o}} = \min_k X_k^s + \left(\frac{\max_k X_k^s - \min_k X_k^s}{\max_k X_k^o - \min_k X_k^o} \right) \left[X_p^o - \min_k X_k^o \right] .$$

Then the distance metric D_t quantifying the dissimilarity between the normalized object vector $X^{\bar{o}}$ and reference scallop vector X^s is given by

$$D_t = \sqrt{\sum_{k=1}^{n} \frac{\|X_k^{\bar{o}} - X_k^s\|^2}{\sigma_k}} \ ,$$

where σ_k refers to the standard deviation of feature k in the reference scallop profile obtained from the look-up table.

To enhance the robustness of the classification layer to small errors in segmentation, nine different masks are used, each centered slightly off the center of the template area. (see Figure 11). This results in nine different feature points, and therefore nine values for the distance metric D_t: say $D_t^{o_1} \ldots D_t^{o_9}$. The distance metric D_{obj} then used for decision is the smallest of the nine: $D_{\mathsf{obj}} = \min_{p \in \{1,\ldots,9\}} D_t^{o_p}$. If D_{obj} is found to be less than a constant



(a) (b)

Figure 12: (a) Precision-Recall curve with D_{thresh} shown as a vertical line; (b) Histogram of template match of segmented scallop objects.

value D_{thresh}, the corresponding object is classified as a scallop.

The classification layer used a template match threshold value $D_{\mathsf{thresh}} = 7$, justified by Figures 12(a)–12(b). The precision-recall curve (*Recall* refers to the fraction of relevant instances identified: fraction of scallops detected over all ground truth scallops; *precision* is the fraction of the instances returned that are really relevant compared to all instances returned: fraction of true scallops over all objects identified as scallops) in Figure 12(a) suggests that the chosen threshold value achieves a recall rate of 97%; for that high detection rate, however, the price to pay is a high chance of false positives as indicated by the low value read off the tail of the precision curve. In the histogram of Figure 12(b), it is seen that for the selected threshold value, the vast majority of scallop objects segmented are actually going to be passed through the classification layer after matching.

Assessment

This multi-layered detection approach was tested on two separate datasets containing 1 299, and 8 049 images respectively. The results are shown in Table 4. As ground truth, only

Table 4: Results of multi-layer scallop classification

	Dataset 1	Dataset 2
Number of images	1,299	8,049
Ground Truth Scallops	363	3,698
Valid Ground Truth Scallops	250	2,781
After Visual Attention Layer	231 (92.4%)	2,397 (86.2%)
After Segmentation Layer	185 (74%)	1,807 (64%)
After Classification Layer	183 (73%)	1,759 (63.2%)
False Positives	17,785	52,456

scallops that were at least 80 pixels horizontally and 60 pixels vertically away from the image boundaries were used. Scallops that were closer to the image boundaries were excluded as they were affected by severe vignetting effects caused by the strobe light on the auv; the boundaries become too dark (see Figure 3) to correct with standard vignetting correction algorithms. In addition, for scallops appearing near the image boundaries the orientation of the characteristic crescents are such that they blend in the dark image boundaries (see Figure 9(a)), and as a result, almost every object in the background in those locations would be mistakenly matched to a scallop. Manual counts performed in (Walker 2013) also used a similar criteria to exclude scallops that were only partially visible near the image boundaries.

Table 4 shows the number of scallops that filter through each layer of the reported approach, and the respective percentage with respect to the number of (away from boundary) valid ground truth scallops (row 3 of Table 4) in the datasets. In dataset 1, which contains 1 299 images, the three-layer filtering results in a final 73% overall recall rate, while in dataset 2 that contains 8 049 images the overall recall rate is 63.2%. At this time it is still unclear what exactly resulted in the higher recall rate in the smaller dataset.

To verify the effectiveness of the classification layer (Layer III) which depends on a customized template matching method, it was compared with a Support Vector Machine (svm) classifier that used a linear kernel. This svm was trained on the segmented objects that were obtained from the segmentation layer (Layer II). This classifier was tested on the dataset containing 8 049 images (same dataset as seen Table 4) and it was found that the total number of scallops detected dropped to 48.5% (compared to 63.2% in our method). However the svm classifier was more effective in decreasing the false positives by roughly 3 times. Finally, the template matching was favored over the svm classifier, because the priority in this work was to maximize the detection rate knowing that at this stage of development, some subsequent manual processing will anyway be necessary. In other words, this implementation leans towards maximizing true positives even at the expense of a large number of false positives.

Discussion

The three-layer automated scallop detection approach discussed here works on feature-poor, low-light imagery and yields overall detection rates in the range of 60–75% . At this juncture, it is important to consider and compare with other available reported scallop detection methods (Einar Óli Guòmundsson 2012; Dawkins et al. 2013) and draw any notable differences between them and the work presented here.

In related work on scallop detection using underwater imaging (Dawkins et al. 2013), reported detection rates are higher, however one needs to stress that the initial imagery data is very different. Specifically, the data sets on which the algorithms (Dawkins et al. 2013) (see also (Dawkins 2011)) operated on exhibit much more uniform lighting conditions, and higher resolution, brightness, contrast, and color variance between scallops and background (see Figure 13). For instance, the color variation between the scallops and background data can be observed by comparing the saturation histogram shown in Figure 13. The histogram of scallop regions in our dataset is often identical to the global histogram of the image, or in other words, the background. On the other hand, the bimodal nature of the saturation histogram of scallop regions in the Woods Hole dataset (Figure 13(c)) makes it easier to separate the foreground from the background.

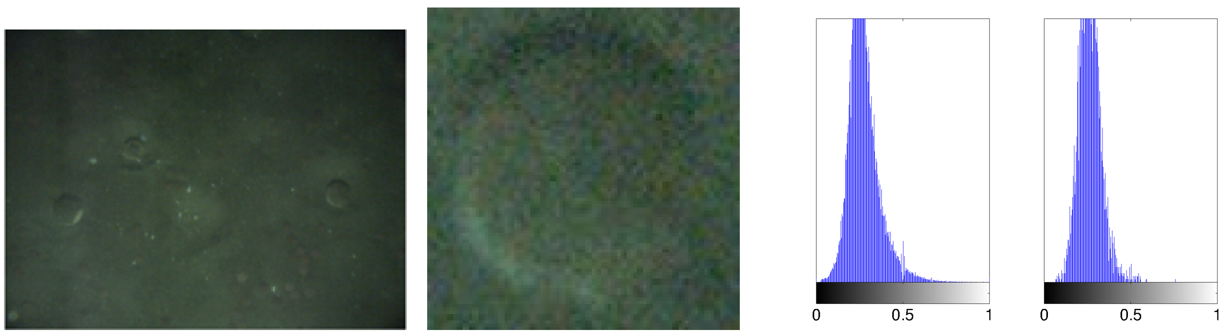
The striking differences between the nature and quality of imagery datasets in these two cases render the results technically incomparable. In particular, the detection algorithm reported in this paper relies heavily on the effect of auv strobe lighting on the collected images, that lets scallops cast characteristic shadows which are subsequently used as features. In contrast, such shadows do not appear around the scallops in the dataset of alternative approaches (Dawkins et al. 2013), because of the different lighting configuration of the HabCam system. In principle, however, with appropriate adjustment of the third layer of the reported algorithm (specifically, selecting different features based on the distinctive characteristics that the particular dataset offers) the approach described here can be adapted to be applicable to datasets collected using very different hardware.

There are advantages in the reported approach compared to the scallop detection framework that uses a series of bounding boxes to cover the entire image (Einar Óli Guòmundsson 2012). The approach of this paper uses just 10 windows per image (as given by tdva), narrowing down the search space much faster. Although the choice of negative instances for the svm classifier of Einar Óli Guòmundsson (2012) still needs to be clarified, our reported classification layer can outperform an alternative svm in terms of detection rates. One should use caution when comparing with the detection rates of Einar Óli Guòmundsson (2012), since these were derived from a select dataset of 20 images and it is also not clear how they would generalize to larger datasets.

Comments and Recommendations

This work is a first step toward the development of an automated procedure for scallop detection, classification and counting, based on low resolution imagery data of the population

(a) (b) (c) Histogram of saturation values of background (left) and cropped scallop (right) from dataset in Dawkins et al. (2013)



(d) (e) (f) Histogram of saturation values of background (left) and cropped scallop (right) from our dataset

Figure 13: Representative samples of different imagery data on which scallop detection algorithms may be called to operate on. Figures 13(a) and 13(d), show an image containing a single scallop from the dataset used by Dawkins et al. (2013) (used with permission from the authors) and the datasets used in this paper respectively. A magnified view of a scallop cropped from Figure 13(a) and 13(d) can be seen in Figures 13(b) and 13(e) respectively. Figure 13(c) gives the saturation histogram of background or the complete image in Figure 13(a) to left and saturation histogram of Figure 13(b) to the right. Similarly, Figure 13(f) gives the saturation histogram of Figure 13(d) to the left and saturation histogram of Figure 13(e) to the right. The bimodal nature of the scallop histogram in Figure 13(c) derived from the dataset used in (Dawkins et al. 2013), clearly portrays the distinguishing appearance of the scallop pixels from the rest of the image, making it easily identifiable. The datasets we used did not exhibit any such characteristics (as seen in Figure 13(f)) to aid the identification of scallops.

partially concealed in its natural environment, and specifically under poor lighting and low contrast conditions. Under such conditions, the performance figures reported are deemed encouraging, but by no means perfect, and there is still room for further improvement. Compared to existing work along this direction, the approach reported in this paper can handle imagery data of much lower quality, and has potential for computational time savings, due to the targeted processing of image regions indicated by visual attention algorithms.

Significant improvements in terms of detection and classification accuracy can be expected is in the context of pre-filtering and processing of raw image data. In the current auv setup, limited onboard memory availability makes it difficult to save raw image data and hence the images are compressed to JPEG format before being saved (raw images are much larger in size and contain more color and light information than compressed JPEG images). Some degree of light, color, and distortion correction (Dawkins et al. 2013) on the raw images before compression will improve classification results, particularly within the segmentation and template matching stages. Another possibility for improvement could be in the direction of reducing the number of false positives. There is a natural trade-off between the template matching threshold and the number of false positives which will penalize detection rates if the former is chosen too low. A specific idea to be explored, therefore, is that of cross-referencing the regions in which include positives against the original, pre-filtered data. These ideas are topics of ongoing and future work.

In this implementation, generic off-the-shelf components for segmentation and template matching were used along with some novel problem-specific realization choices. Although there exist some low-level technical challenges associated with these component integration, there is also room for improvement in the implementation of these components themselves, in terms of computational efficiency. In the current implementation, the graph-cut based image segmentation component is taxing in terms of computation time, and this area is where computational improvements are likely to yield the largest pay-off. On the other hand, the overall architecture is modular, in the sense that the segmentation and classification layers of the procedure could in principle be implemented using a method of choice, once appropriately interfaced with the neighboring layers and due to the fact that it allows retraining for other object detection problems with very different backgrounds or characteristic object features.

References

Basseville, M. and Nikiforov, I. V. (1993). *Detection of abrupt changes: theory and application*, volume 104. Prentice Hall.

Caddy, J. (1975). Spatial model for an exploited shellfish population, and its application to the georges bank scallop fishery. *Journal of the Fisheries Board of Canada*, 32(8):1305–1328.

Chernov, N. (2010). *Circular and linear regression: Fitting circles and lines by least squares.* Taylor and Francis.

Dawkins, M. (2011). Scallop detection in multiple maritime environments. Master's thesis, Rensselaer Polytechnic Institute.

Dawkins, M., Stewart, C., Gallager, S., and York, A. (2013). Automatic scallop detection in benthic environments. In *IEEE Workshop on Applications of Computer Vision*, pages 160–167.

Edgington, D. R., Cline, D. E., Davis, D., Kerkez, I., and Mariette, J. (2006). Detecting, tracking and classifying animals in underwater video. In *OCEANS 2006*, pages 1–5.

Einar Óli Guòmundsson (2012). Detecting scallops in images from an auv. Master's thesis, University of Iceland.

Enomoto, K., Masashi, T., and Kuwahara, Y. (2010). Extraction method of scallop area in gravel seabed images for fishery investigation. *IEICE Transactions on Information and Systems*, 93(7):1754–1760.

Enomoto, K., Toda, M., and Kuwahara, Y. (2009). Scallop detection from sand-seabed images for fishery investigation. In *2nd International Congress on Image and Signal Processing*, pages 1–5.

Fearn, R., Williams, R., Cameron-Jones, M., Harrington, J., and Semmens, J. (2007). Automated intelligent abundance analysis of scallop survey video footage. *AI 2007: Advances in Artificial Intelligence*, pages 549–558.

Fisheries of the United States (2012). Fisheries of the United States, Silver Spring, MD. Technical report, National Marine Fisheries Service Office of Science and Technology.

Forrest, A., Wittmann, M., Schmidt, V., Raineault, N., Hamilton, A., Pike, W., Schladow, S., Reuter, J., Laval, B., and Trembanis, A. (2012). Quantitative assessment of invasive species in lacustrine environments through benthic imagery analysis. *Limnology and Oceanography: Methods*, 10:65–74.

Gallager, S., Singh, H., Tiwari, S., Howland, J., Rago, P., Overholtz, W., Taylor, R., and Vine, N. (2005). High resolution underwater imaging and image processing for identifying essential fish habitat. In Somerton, D. and Glentdill, C., editors, *Report of the National Marine Fisheries Service Workshop on Underwater Video analysis*, NOAA Technical Memorandum NMFS-F/SPO-68, pages 44–54.

Hart, D. R. and Rago, P. J. (2006). Long-term dynamics of US Atlantic sea scallop Placopecten magellanicus populations. *North American Journal of Fisheries Management*, 26(2):490–501.

Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259.

Jenkins, S., Beukers-Stewart, B., and Brand, A. (2001). Impact of scallop dredging on benthic megafauna: a comparison of damage levels in captured and non-captured organisms. *Marine Ecology Progress Series*, 215:297–301.

Koch, C. and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4(4):219–227.

McGavigan, C. (2012). A quantitative method for sampling littoral zooplankton in lakes: The active tube. *Limnology and Oceanography: Methods*, 10:289–295.

Naidu, K. and Robert, G. (2006). Fisheries sea scallop placopecten magellanicus. *Developments in Aquaculture and Fisheries Science*, 35:869–905.

National Marine Fisheries Service Northeast Fisheries Science Center (NEFSC) (2010). 50th northeast regional stock assessment workshop (50th saw) assessment report. Technical Report 10-17, p.844, US Dept Commerce, Northeast Fisheries Science Center.

Navalpakkam, V. and Itti, L. (2006). An integrated model of top-down and bottom-up attention for optimizing detection speed. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 2049–2056.

Oremland, L., Hart, D., Jacobson, L., Gallager, S., York, A., Taylor, R., and Vine, N. (2008). Sea scallop surveys in the 21st century: Could advanced optical technologies ultimately replace the dredge-based survey? Presentation made to the NOAA Office of Science and Technology.

Poor, H. V. and Hadjiliadis, O. (2009). *Quickest detection*. Cambridge University Press.

Rosenberg, A. A. (2003). Managing to the margins: the overexploitation of fisheries. *Frontiers in Ecology and the Environment*, 1(2):102–106.

Rosenkranz, G. E., Gallager, S. M., Shepard, R. W., and Blakeslee, M. (2008). Development of a high-speed, megapixel benthic imaging system for coastal fisheries research in alaska. *Fisheries Research*, 92(2):340–344.

Serchuk, F., Wood, P., Posgay, J., and Brown, B. (1979). Assessment and status of sea scallop (placopecten magellanicus) populations off the northeast coast of the united states. In *Proceedings of the National Shellfisheries Association*, volume 69, pages 161–191.

Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905.

Spampinato, C., Chen-Burger, Y.-H., Nadarajan, G., and Fisher, R. B. (2008). Detecting, tracking and counting fish in low quality unconstrained underwater videos. In *3rd International Conference on Computer Vision Theory and Applications*, pages 514–519.

Stelzer, C. P. (2009). Automated system for sampling, counting, and biological analysis of rotifer populations. *Limnology and oceanography: Methods*, 7:856.

Taubin, G. (1991). Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(11):1115–1138.

Trembanis, A. C., Phoel, W. C., Walker, J. H., Ochse, A., and Ochse, K. (2011). A demonstration sea scallop survey of the federal inshore areas of the new york bight using a camera mounted autonomous underwater vehicle. Technical report, National Oceanographic and Atmospheric Administration.

Walker, J. (2013). Abundance and size of the sea scallop population in the mid-atlantic bight. Master's thesis, University of Delaware.

Walther, D. and Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, 19(9):1395–1407.

Williams, R., Lambert, T., Kelsall, A., and Pauly, T. (2006). Detecting marine animals in underwater video: Let's start with salmon. In *Americas Conference on Information Systems*, volume 1, pages 1482–1490.