

Automated Detection of Scallops in their Natural Environment

Prasanna Kannappan and Herbert G. Tanner

Abstract—Automating the counting of marine animals like scallops benefits marine population survey efforts. These surveys are tools for policy makers to regulate fishing activities, and sources of information for biologists and marine ecologists interested in population statistics of marine species. In this paper we discuss some practical difficulties that arise in the scallop detection problem from visual data, and propose a solution based on top-down visual attention. We assess the performance of the proposed method against a comparable and related method which has recently been employed in literature, using a significant amount of ground truth data.

Keywords—Scallop identification; visual attention; marine surveys.

I. INTRODUCTION

Counting marine animals is of interest to groups of biologists and marine ecologists as well as to fishery agencies, both for scientific and commercial purposes. Besides providing insights to research problems in marine biology, these surveys also aid policy makers in regulating recreational and commercial exploitation of these resources. This paper reports on an algorithmic method to count sea animals like scallops automatically from seabed images.

Such automation is needed because traditional dredge-based methods are very invasive and do not provide accurate estimates of population density. In these methods, the seabed is dredged scooping up a strip of the ocean floor, and the marine animals found in the sediment are counted manually before they are thrown back to the sea. The survey numbers produced this way are then generalized to give an estimate of the species abundance in the particular area. Advances in underwater vehicle technology have enabled the use of an autonomous underwater vehicle (AUV) or a remotely operated vehicle (ROV) for this purpose. The underwater vehicles use onboard cameras to collect video footage and allow a species count from this data in a non-invasive manner [1]. These methods also have an added advantage of providing ways to visualize the habitat of the marine animals of interest without disturbing it, and relate population growth to specific substrates. Despite these technological advancements in the domain of marine population data collection, counting of animals is still done manually through arduous observation of hours of video footage. Robotic-assisted marine species surveys will therefore benefit from the automation of the counting process.

There have been attempts to counting marine species using stationary underwater cameras [2], [3]. In this general framework, salmon are counted through background subtraction

and shape detection [4]. 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 addition, this application requires the use of mobile underwater vehicles, rather than stationary mooring cameras.

One way to approach the problem of detecting marine animals from seabed images is by identifying the points in the image which are most likely to contain objects, as those 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 dataset where there is a sudden change in the underlying distribution from which the data is generated. Formal mathematical approaches to determining this change in distribution are discussed in [5] and [6]. The challenge there is that some prior knowledge about the underlying distribution and / or the change time is required. Without several technical assumptions, sometimes of debatable validity in the specific application context, modeling the background distribution from image data can be problematic.

For scallops in particular, the challenges in automating the identification process based on visual data come primarily from the absence of features on these creatures that would clearly distinguish them from their natural environment. In addition, collecting visual data from the species' natural habitat yields images with a wide range of illumination variations, presence of significant amount of speckle noise, sensor noise due motion of underwater platforms, and poor resolution. Some of the scallops are also partially, or almost completely covered by sediment, obscuring the scallop shell features. To overcome these impediments, a highly robust detection mechanism is required.

Some automation attempts to count scallops in artificial environments can be found in [7], [8]. Both approaches employ a detection mechanism based on intricate distinguishing features like fluted patterns in scallop shells and exposed shell rim. Though this might be effective in artificial scallop beds with stationary cameras and minimal sensor noise, it is difficult to identify these patterns in scallops present in their natural environment, especially in images of poor resolution taken several meters away from the target—it might not be advisable to drive an underwater vehicle too close to the

The authors are with the Department of Mechanical Engineering, University of Delaware. Email: {prasanna, btanner}@udel.edu

This work is supported by NSF under RI grant #0913015.

ocean floor.

There is some work on scallop detection in their natural environment [9]. In this work, an elaborate machine learning-based approach is discussed. Though real seabed images are used for analysis, the data sets to test the efficiency of the method are relatively small, in the order of 30 images. From this study alone, it is not clear if such a method can be used effectively in cases of large data sets comprising several thousand seabed images, collected from actual AUV missions. An interesting example of another machine-learning method applied to the problem of scallop detection is [10]. It utilizes the concept of bottom-up visual attention (BUVA). The approach is promising but interestingly, 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.

What is done in this paper is the application of top-down visual attention (TDVA) to the problem of identification of scallops in their natural environment. The difference to the earlier bottom-up approach is that now the method is biased to pick up candidate targets that have features that are commonly found in scallops. To validate the proposed method, we assess the top-down approach side-by-side with an implementation of the bottom-up approach, on the same set of data. In addition to the application of top-down visual attention to this problem being the first reported in literature, this paper enhances the standard approach [11] with some operational novelties, and also uses ground-truth measurements for the validation of the method’s performance, over a large set of visual data.

The paper is structured as follows: Section II describes what we mean by detecting a scallop, it presents some candidate scallop features, and describes the challenges associated with the problem. Section III reviews briefly the method of visual attention, and Section IV describes our specific setting, data, and assumptions. Section V discusses implementation details related to top-down visual attention, and Section VI comments on the results from the application of the top-down and bottom-up visual attention variants. Section VII summarizes the paper and outlines future research directions.

II. PROBLEM DESCRIPTION AND CHALLENGES

In more technical terms, the problem addressed in this paper is that of detecting candidate regions in an image which have a high probability of containing a scallop. These regions can then be further analyzed to determine if they indeed contain scallops, but this further analysis is not part of this paper.

Several challenges have to be overcome in solving this problem. First, the high levels of speckle noise evident in Figs. 1–2 make basic morphological filtering operations, like thresholding and edge detection, problematic. In addition, the edges of scallops are often not very distinct, they may be



Fig. 1. Seabed image with scallops shown by red circles

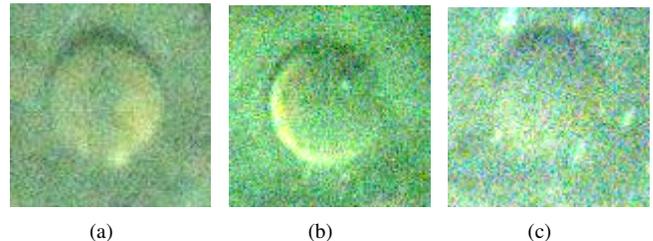


Fig. 2. (a) Scallop with yellowish tinge and dark crescent (b) Scallop with yellowish tinge and bright shell rim crescent (c) Scallop with no prominent crescents and texturally identical to the background

discontinuous, and often gradually dissolve into the background (see Fig. 2(b)). One of the reasons for this is the angle of the AUV strobe light, which tends to strengthen the crescents to the top of the scallop and weaken the ones on the lower end. Furthermore, there are other creatures that can appear in the vicinity of scallops—like sand dollars—which are morphologically similar and can affect the visible profile of scallops. In several cases, speckle noise reduction filters can weaken the edges and exacerbate the problem. Finally, scallops can be partially or completely buried by sediment, making them texturally identical to the background (see Fig. 2(c)). As a result, most of the scallop area is segmented as background, leaving any detection algorithm purely dependent on weak edges or crescents at the periphery of the scallop.

III. PRELIMINARIES: VISUAL ATTENTION

A biologically inspired approach to addressing the problem of detecting points of interest in an image is to consider a phenomenological model of the human visual system, which tries to capture our inbuilt ability to filter in interesting details from images. The robustness of this mechanism to noise, and its ability to process large volume of information is justified through several competing neuro-biological models, on which the concept of *visual attention* is based on. Visual attention is a theory around a hypothesis according to which the human visual system preferentially processes certain parts of an image first. There are two variants in visual attention, bottom-up (BUVA) and top-down (TDVA) [12]. The bottom-up variant picks out the regions in an image

which are sufficiently discriminative with respect to their surroundings. Sometimes the visual attention mechanism is preconditioned, and the search is directed toward objects with certain attributes like a specific color or texture. The result of implementing this preconditioning gives rise to top-down visual attention (TDVA) [12].

A. Bottom-Up Visual Attention

Bottom-up visual attention can be used to pinpoint regions in an image where the value of some pixel attributes may be a sign of a change with respect to the rest of the image. In contrast to mathematical formulations of the problem of detecting changes in distribution [6], here the decision is purely heuristic. 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 *saliency*. Points with high saliency values are processed first.

According to this hypothesis [13], 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 so-called *center-surround* feature maps. 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 based on their saliency values.

A computational model for the above visual attention mechanism is proposed in [11]. 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 ($\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$). The image information in each sub-stream is scaled in eight different factors, $2, \dots, 9$; for each factor k the image data is essentially scaled to size $\frac{1}{2^k}$, resulting in some loss of information as the scale increases. The resulting image data for each factor constitutes the *spatial scale* for the particular sub-stream.

What happens then is that the sub-stream feature information are resized and compared, but because of the loss of information and the interpolation taking place during resizing, the resized feature maps do not match exactly across different scales, even when brought down to the same size. The *center-surround* operator \ominus takes pixel-wise differences between resized sub-streams and exposes those mismatches. For the intensity stream, we would write:

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

where c and s are indices for two different spatial scales with $c \in \{2, 3, 4\}$, $s = c + \delta$, for $\delta \in \{3, 4\}$. Center-surround feature maps are computed for each sub-stream in color and orientation streams in the same way.

With a total of seven sub-streams (two for color, one for intensity and four for orientation), we get a total of 42 center-surround feature maps. The center-surround feature maps for

each original stream (color, intensity, and orientation) are then combined back into three *conspicuity maps*: one for color \bar{C} , one for intensity \bar{I} , and one for orientation \bar{O} . The intensity conspicuity map, for example, is given as

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

where now the \oplus cross-scale operator works in a fashion similar to \ominus , with the difference being that data in resized maps from different scales is now pixel-wise added. In (2) we see the map normalization operator $\mathcal{N}(\cdot)$, which basically scales a whole map by multiplying the values in its argument with 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. Conspicuity maps are then simply combined—no cross-scale operation is needed here—to get the *saliency map*:

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

where $w_{\bar{k}}$ is in general a user-selected stream-specific weight. In 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 [11], [14] to compute the maxima—other methods are of course possible. When a local maximum is found, the focus of attention is supposedly shifted to this point, and the point becomes a visual fixation.

B. Top-Down Visual Attention

In a top-down visual attention approach, fixations are biased toward specific targets of interest. This is done by adjusting the weights in (3) and (2) to favor features unique to targets over other image differences. The learning procedure in [15] can suggest weights that maximize the signal-to-noise ratio, in the form

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

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

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

where N is the number of images in the learning set, N_{rT} and N_{rD} are the number of targets (scallop) and distractors (similar objects) in the r -th learning image, P_{uvT_z} is the mean saliency 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).

IV. DATA ACQUISITION

The data used in this paper was collected as part of the scallop survey reported in [1]. In this survey, a Gavia AUV with an onboard Point Grey Scorpion camera model 20S0 mounted on the nose of the vehicle was used to collect over 250 000 images of the seabed along the New York – New Jersey coast. The images were then manually annotated, and a scallop count was performed on them. The annotation

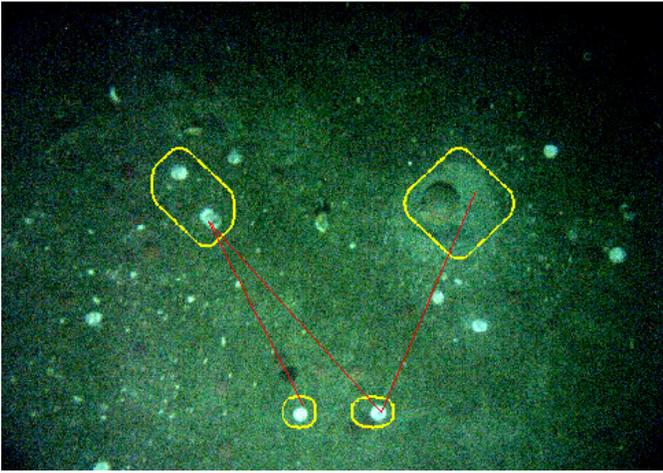


Fig. 3. Illustration of fixations. The red lines indicate the order in which the fixations were detected with the lower-left fixation being the first. The yellow outline is the proto-object around the fixation.

introduced, contained information about the center and radius of the scallops identified, assuming them to be roughly circular. The images were of resolution 800×600 pixels, and were captured from an height of 2 m above sea floor, with a horizontal viewing angle of 49.92° . Calculations based on the viewing angle and altitude from the sea floor indicate that the average area covered per image is $1.86 \text{ m} \times 1.40 \text{ m}$ of sea floor, with an approximate resolution of 2 mm/pixel. The strobe light source attached to the rear of the AUV resulted in scallops presenting a characteristic dark crescent shadow (dark crescent at the top of the scallop). Some prominent sources of noise observed were in the form of speckle noise and vignetting effects caused by the strobe lights on the AUV.

V. METHODOLOGY

The selection of features on scallops is an issue that can be open to debate, and different reasonable suggestions can be offered depending on context. In our data set, (see Fig. 1 for a representative sample) we could not distinguish some unequivocal feature choices, but could nevertheless identify some recurring patterns.

One of these patterns is a dark crescent on the upper perimeter of the scallop shell, which is in fact the shadow against the AUV strobe light, cast by the open upper scallop shell opposite to the hinge (see Fig. 2(a)). Another pattern that could serve as a scallop feature in this data set is a frequently occurring bright crescent on the periphery of the scallop, generally being the visible edge of the scallop shell, as the center of the shell is often covered by sediment (see Fig. 2(b)). Yet another recurring pattern is a yellowish tinge associated with the composition of the scallop image (see Fig. 2(b)).

In this paper we apply TDVA to the problem of scallop detection, that is, we bias the weights in (2)–(3) according to (4) for the purpose of skewing the appearance of saliency peaks toward regions with scallops as opposed to other “potentially interesting” sea-floor objects. As a part of the

procedure to learn the top-down saliency weights, bottom-up saliency computation is performed first on 243 annotated images collectively containing 300 scallops. An example of bottom-up computation on an image is illustrated in Fig. 4. The figure portrays the process of computing the color, intensity, and orientation conspicuity maps from the original image (as described in Section III-A). These conspicuity maps are combined to compute the saliency map. The intermediate step of computing the center-surround feature maps has been omitted from the figure for the sake of clarity.

Fixations are then identified in the saliency maps, and each fixation marks a region of interest in each image. Figure 3 shows an image with the top (i.e., the ones with the highest saliency value) four fixations, among which one is a scallop. The yellow outline around the fixation is what is called the *proto-object* [11] of that fixation. Experience suggested that proto-objects rarely contain targets. They are usually just regions texturally identical to the fixation identified. This can be further explained by the fact that in most cases, the center of the scallop is texturally identical to the background, and thus the crescents are the only salient features truly associated with the scallop.

We use a rectangular window of size 100×100 centered around fixations, and if a scallop center lies within this window, the corresponding fixation is labeled as a *target fixation*. If there is no scallop present in this window, it is labeled as a *distractor fixation*. The diameter of the scallops in our dataset varied from 20 to 70 pixels. This factor combined with our empirical observation that fixations tend to occur close to the periphery of a scallop resulted in choosing a window size of 100×100 which in most cases contained the center of the scallop. We found empirically that ten fixations are usually sufficient to capture almost all scallops in any given image. Furthermore, the average number of scallops per image turned out to be less than two. Taking ten fixations accommodated even rare cases where there were more than five scallops in the same image. In none of the images considered were there more than ten scallops.

In each one of these processed images in the learning set, we determine the target and distractor regions in the feature and conspicuity maps. 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 are used to compute the top-down weights for feature maps and conspicuity maps, respectively, using (4).

The resulting top-down conspicuity map weights w_T , w_C , w_O are shown in Table II. The set of feature map weights for each center-surround scale w_{cs} for every feature is listed in Table I. Once the top-down weights are computed, specialized feature maps and saliency maps that favor detection of scallops over just any other interesting feature can be computed, by using the computed weights in (2) and (3) respectively.

On the saliency maps of our implementation of TDVA

TABLE II
TOP-DOWN WEIGHTS FOR CONSPICUITY MAPS

Color	Intensity	Orientation
1.4354	1.1644	0.4001

TABLE III
COMPARISON OF BUVA AND TDVA

	BUVA	TDVA
Number of scallops detected	325	327
Percentage of scallops detected	89.8	90.3
Fixations required to detect first scallop	2.51	2.48
Fixations required to detect all scallops	2.86	2.92

we use dynamic thresholds. This is done by controlling the convergence time required for the winner-takes-all neural network. When convergence requires more than 10000 iterations, it is highly unlikely for the particular fixations to be relevant. In principle, even specks of noise can produce fixations if the particular neural network is allowed to evolve indefinitely. 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.

VI. RESULTS AND DISCUSSION

A. Validation Procedure

To assess the performance of our top-down visual attention based scallop detector, we implement a bottom-up visual attention (BUVA) algorithm, along the lines of [10], and compare the results side-by-side. The methodology for implementing BUVA is described in Section V; the difference is primarily in the selection of the weights: uniform weights equal to 1 for BUVA as opposed to the custom weights of Tables I and II for TDVA.

B. Performance Analysis

A set of 2388 images is used to test the performance of the two visual attention variants. The dataset contained 362 labeled scallops. Table III summarizes the performance of top-down and bottom-up visual attention on the scallop detection problem. The top-down approach (TDVA) performs slightly better than bottom-up (BUVA), detecting 90.3% of the scallop population, compared to 89.8% of BUVA. The average number of fixations required to detect the first scallop by bottom-up is slightly less than top-down. The same can be said for the performance of bottom-up over top-down in average number of fixations required to detect all scallops in the images.

C. Discussion

Conceptually, top-down visual attention is designed to perform better than bottom-up visual attention in detecting specific objects with unique distinguishing characteristics, even in cluttered scenes containing several irrelevant objects [16]. In the case considered in this paper, however, the object

of interest (a scallop) does not have many distinguishing characteristics to bank on. Some features, like the yellowish tinge, or the dark crescents, are not common to all scallops. Despite wide range of differences between scallop images, those which show prominently visible features bias the top-down weights towards these specific features. The yellowish tinge in some scallops result in the blue-yellow color feature maps being weighted above red-green feature maps which can be seen in Table I. Similar remarks can be made with reference to Table II, where the orientation conspicuity map has the least weight, since scallops are uniformly circular in shape and thus the orientation features do not fit into any of the 4 orientation sub-streams. A good number of scallops are visually salient but devoid of characteristic scallop features. These scallops do not help when used in the learning set. A top-down approach tuned for the scallop-like features discussed in Section V will either need considerably more fixations to identify a scallop devoid of these features, or just miss it completely. On the other hand, a bottom-up approach which works purely on detecting points merely different from background behaves without bias, and detects any object which stands out, irrespectively of whether it is a scallop or not.

Despite these limitations, and in the presence of significant amount of speckle noise and illumination variations in images, the TDVA method still performs slightly better than its BUVA counterpart, detecting approximately 0.5% more scallops. Overall, when it comes to object detection applications where the targets do not have many distinguishing features, it appears that there is a slight performance improvement when using TDVA compared to BUVA. Whether this performance enhancement justifies the additional computational overhead involved in training and tuning the weights of TDVA eventually depends on the particular application setting.

VII. CONCLUSION AND FUTURE WORK

For object detection applications involving significant amount of noise in the data, and with minimal distinguishing object features, top-down visual attention appears to be an effective tool, and yields slightly better performance compared to its bottom-up counterpart. This hypothesis is tested on a large data set of images of scallops in their natural habitat, taken by an AUV. We postulate that combining one of the variants of visual attention with a robust segmentation and classification algorithm can further improve the observed performance. It is also an interesting possibility to determine the data acquisition parameters like camera angle and AUV height from the seafloor that optimize the performance of our method.

ACKNOWLEDGEMENTS

The authors would like to thank Art Trembanis and Justin Walker for the data used in this paper which was collected as a part of their scallop survey [1].

TABLE I
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

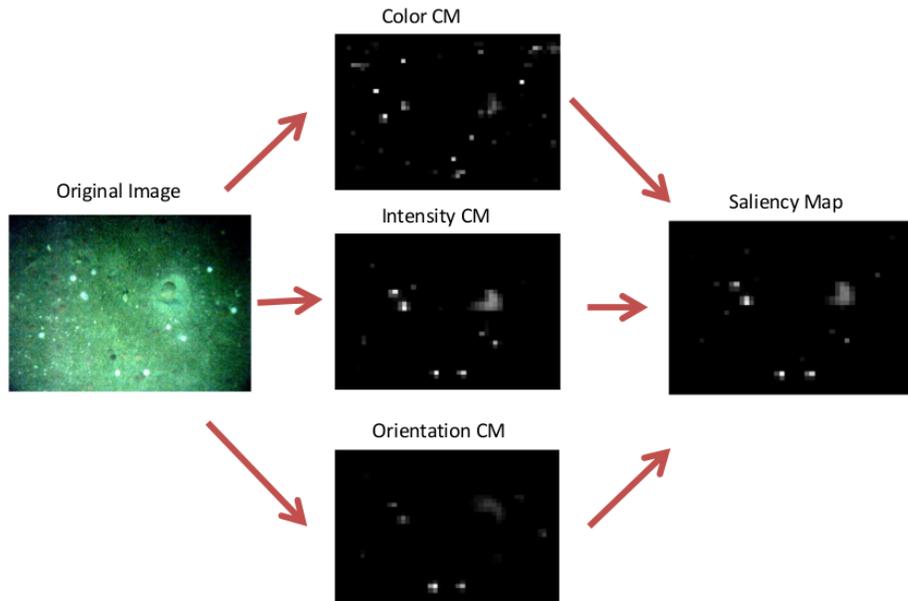


Fig. 4. Illustration of saliency map computation

REFERENCES

- [1] A. C. Trembanis, W. C. Phoel, J. H. Walker, A. Ochse, and K. Ochse, "A demonstration sea scallop survey of the federal inshore areas of the new york bight using a camera mounted autonomous underwater vehicle," National Oceanographic and Atmospheric Administration, Tech. Rep., 2011.
- [2] C. Spampinato, Y.-H. Chen-Burger, G. Nadarajan, and R. B. Fisher, "Detecting, tracking and counting fish in low quality unconstrained underwater videos," in *3rd International Conference on Computer Vision Theory and Applications*. Citeseer, 2008, pp. 514–519.
- [3] D. R. Edgington, D. E. Cline, D. Davis, I. Kerkez, and J. Mariette, "Detecting, tracking and classifying animals in underwater video," in *Oceans'06 MTS/IEEE-Boston Conference and Exhibition*. IEEE, 2006, pp. 1–5.
- [4] R. Williams, T. Lambert, A. Kelsall, and T. Pauly, *Detecting Marine Animals in Underwater Video: Let's Start with Salmon.*, 2006.
- [5] M. Basseville and I. V. Nikiforov, *Detection of abrupt changes: theory and application*. Prentice Hall Englewood Cliffs, NJ, 1993, vol. 104.
- [6] H. V. Poor and O. Hadjiladis, *Quickest detection*. Cambridge University Press, 2009.
- [7] K. Enomoto, M. Toda, and Y. Kuwahara, "Scallop detection from sand-seabed images for fishery investigation," in *2nd International Congress on Image and Signal Processing*. IEEE, 2009, pp. 1–5.
- [8] K. Enomoto, T. Masashi, and Y. Kuwahara, "Extraction method of scallop area in gravel seabed images for fishery investigation," *IEICE Transactions on Information and Systems*, vol. 93, no. 7, pp. 1754–1760, 2010.
- [9] M. Dawkins, "Scallop detection in multiple maritime environments," Master's thesis, Rensselaer Polytechnic Institute, 2011.
- [10] R. Fearn, R. Williams, M. Cameron-Jones, J. Harrington, and J. Semmens, "Automated intelligent abundance analysis of scallop survey video footage," *AI 2007: Advances in Artificial Intelligence*, pp. 549–558, 2007.
- [11] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [12] R. Desimone and J. Duncan, "Neural mechanisms of selective visual attention," *Annual Review of Neuroscience*, vol. 18, no. 1, pp. 193–222, 1995.
- [13] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, no. 4, pp. 219–227, 1985.
- [14] D. Walther and C. Koch, "Modeling attention to salient proto-objects," *Neural Networks*, vol. 19, no. 9, pp. 1395–1407, 2006.
- [15] V. Navalpakkam and L. Itti, "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*, vol. 2. IEEE, 2006, pp. 2049–2056.
- [16] S. Frintrop, *VOCUS: A visual attention system for object detection and goal-directed search*. Springer, 2006, vol. 3899.