Towards Automated Classification of Seabed Substrates in Underwater Video

Pugh, Matthew; Tiddeman, Bernie; Dee, Hannah Mary; Hughes, Phil

Published in:
Computer Vision for Analysis of Underwater Imagery Workshop

DOI:
10.1109/CVAUI.2014.18

Publication date:
2014

Citation for published version (APA):
Pugh, M., Tiddeman, B., Dee, H. M., & Hughes, P. (2014). Towards Automated Classification of Seabed Substrates in Underwater Video. In *Computer Vision for Analysis of Underwater Imagery Workshop: International Conference on Pattern Recognition* (pp. 9-16). IEEE Press. https://doi.org/10.1109/CVAUI.2014.18

AUTOMATIC CLASSIFICATION OF SEABED SUBSTRATES IN UNDERWATER VIDEO

A. Matthew Pugh, Bernard Tiddeman, Hannah Dee

Department of Computer Science
Aberystwyth University,
Aberystwyth
Wales UK

C. Philip Hughes

Friends of Cardigan Bay
Aberystwyth
Wales UK

ABSTRACT

In this work, we present a system for the automated classification of seabed substrates in underwater video. Classification of seabed substrates traditionally requires manual analysis by a marine biologist, according to an established classification system. Accurate, consistent and robust classification is difficult in underwater video due to varying lighting conditions, turbidity and method of original recording. We have developed a system that uses ground truth data from marine biologists to train and test per-frame classifiers. In this paper we present preliminary results of this using various feature representations (histograms, Gabor wavelets) and classifiers (SCM, kNN).

***Index Terms*—** Underwater Video, Seabed Classification, Substrates, Computer Vision, Texture, Wavelets, Machine Learning

1. INTRODUCTION

Marine habitat monitoring and study has been a subject of interest for research as technology has permitted new approaches to this historically manually-performed task [1, 2, 3]. It directly relates to ongoing ecological surveillance methods, the monitoring of climate change and the management of fisheries. Gaining useful information on subaquatic environments has historically been difficult [3]. Direct-contact monitoring is expensive both in money and time. When considering even specific regions of interest, the sheer scale and area to cover makes this an unfeasible task.

Cardigan Bay is the largest oceanic bay in Wales, located on the western coast with Bardsey Island in the north, and Strumble Head in the west. A local organisation, Friends of Cardigan Bay (FoCB), engage in monitoring the bay's habitat and ecology. We target our approach to the marine habitats in this area, working directly with the organisation's researchers. As with a number of coastal inlets, most of the subaquatic region at Cardigan Bay has gone relatively unexplored, save for specific mandates. We investigate the use of computer vision and machine learning techniques in order to automatically classify seabed ecology. Building upon work in the areas of texture modelling, understanding and representation [1, 4], we evaluate their application to underwater video analysis.

Numerous factors affect the utility of the resultant video and images, both natural and mechanical. The physical properties of sea-water mean that obtaining uniform illumination is difficult, specifically when the depth of the sea increases. Organic particles can obscure frame clarity, and add motion to a scene which can be difficult to cleanly disregard. Colour information varies both within and between videos due to differences in lighting at different depths. Visibility can be substantially affected given the turbidity (the cloudiness of a liquid given particles within, shown in Figure 2) present in the field of view. Finally, limitations of the recording hardware, such as low resolution or low frame-rate capture can also adversely affect any attempts to automatically analyse the video content.

Existing work in this area has investigated the improvement of capturing, cataloging and understanding video data. Most prominently, MBARI's AVED system deals with recognition and classification of different fish species, by identifying regions of interest in a frame through the use of visual saliency, before submitting this area of interest for further analysis [5]. A similar example is the Fish4Knowledge project at The University of Edinburgh, which focused on extracting information from video based on user queries [6, 7]. Research into the effects of climate change on species richness of marine fishes and the optimisation of fishing strategies using these methods continues to be performed. This paper looks at building a more efficient way of selecting regions of video which are known by marine biologists to have a higher probability of containing the sought-after marine species.

The Countryside Council for Wales (CCW) produced a catalog of classifications for marine habitats present in the seas around Wales [8]. In total, 31 distinct classifications are defined, of which 13 have been identified as candidates for the areas covered in our source material, listed in Table 1. The omitted categories have been excluded as marine biologists have confirmed they are not encountered in the surveyed areas.

Table 1. Classes from the CCW schema selected as candidates.

ID	Short Description
0	No relevant data present
14	Vertical subtidal rock with associated community
16	Coarse sands and gravels with communities that include large and/or long lived bivalves
17	Maerl beds
18	Stable predominantly subtidal fine sands
19	Subtidal stable muddy sands, sandy muds and muds
20	Predominantly subtidal rock with low-lying and fast growing faunal turf
22	Shallow subtidal rock with kelp
23	Kelp and seaweed communities on sand scoured rock
24	Dynamic, shallow water fine sands
27	Biogenic reef on sediment and mixed substrata
28	Stable, species rich mixed sediments
29	Unstable cobbles, pebbles, gravels and/or coarse sands supporting relatively robust communities
31	Seagrass beds

Table 2. Our texture-based classification schema.

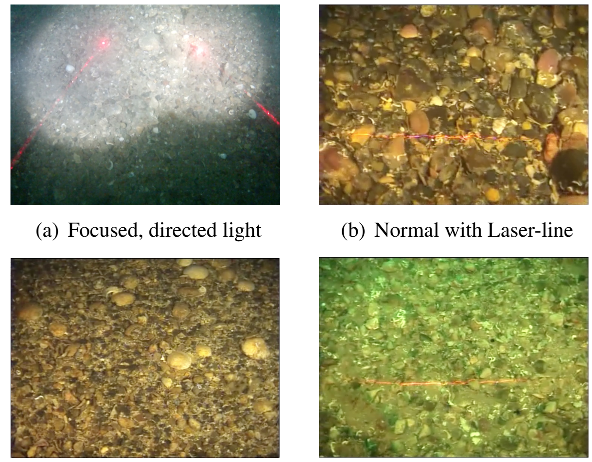
ID	CCW	Aesthetic Description
0	0	No relevant data present
I	18, 19, 24	Fine sands
II	16, 28, 29, 31	Coarse sands with occasional rocks and fauna
III	20	Pebbled seabed with occasional rocks
IV	14	Predominately large-boulders
V	17, 22, 23, 27	Coral & rich in organic life

The CCW schema provides a foundation upon which a more concise schema for classification based on visual properties may be derived. This is necessary as a number of the classifications require sampling the seabed, to disambiguate classes that are very similar visually. The classes in Table 2 map to one or more possible classes in the CCW document, and are visually distinct.

2. METHODS

Ground truth was collected using a custom interface permitting a marine biologist to assign one of the aforementioned classes to ranges of frames in the videos. This range-based markup is illustrated in Figure 3.

A number of different video sets are used in this research, collected via two methods. We focus on two of these sets in this paper, and select videos from each. It is not possible to use every video as many have the same substrate throughout, so here we select videos containing multiple substrate changes.



(a) Focused, directed light (b) Normal with Laser-line

(c) Normal (Desirable) (d) Heavy Green Hue

Fig. 1. Contrast between varying illumination patterns of seabed substrates in source video. The four frames all correlate to class III.

Fig. 2. Typical scene demonstrating difficulties of automated analysis, specifically when underwater conditions include high turbidity.

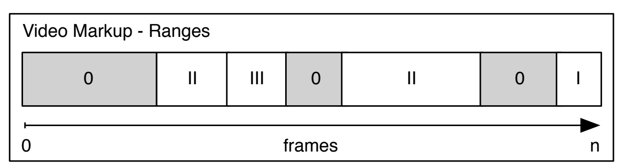


Fig. 3. A sample video is considered a collection of n frames that are split into continuous ranges corresponding with our classification schema, based on manual analysis. Class 0 represents non-viable training / testing frames.

1. Sled with GoPro attached - Launched from a boat and attached via rope, the sled trawls the seabed collecting video data from a GoPro camera, equipped with 2 stationary lights as demonstrated in Figure 1(a). These videos were generously offered by Bangor University collected whilst researching the effects of trawling on the seabed [2]. We evaluate videos from sites 2, 3 &

10.

2. CCTV pin camera - A technique used by FoCB, a commercial DVD recorder system is used with a pin-camera dropped over the boat's edge. This is a directional sensor attached via cable, and sending signal through, to a DVD recorder on the boat's deck. The camera component is attached to a weight to control depth. This method relies on the operator's personal knowledge of local seabed depths and environment. The camera is lowered and raised manually via the boat's deck. This is delivered in PAL interlaced, and pre-processed to remove artefacts before use.

Due to the nature of underwater video captured via trawler, seabed disturbance is to be expected upon the trawler's impact. During the marking up phase, the start and end sections of a classification range are set to omit these frames.

To alleviate problems related to non-uniform illumination and dominant colours in the source video, RGB and greyscale histograms were generated. Colour correction is performed using histogram equalisation. In the case of colour histograms, RGB sub-channel histograms were horizontally-concatenated into a feature vector of the form $\mathbb{R}^{781 \times 1}$. This representation loses colour channel correlation.

Our second approach to the problem is the use of textural image descriptors. This involves the use of a number of Gabor filters defined at a number of different orientations. It has been shown that Gabor filters in this way approximate the cell receptors of the mammalian visual system [9].

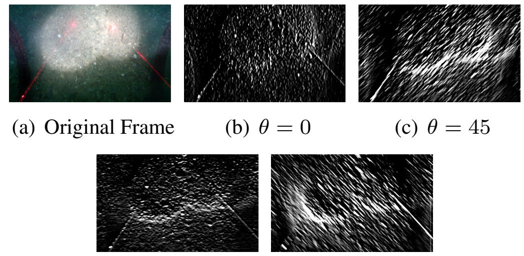
The filter $g(x, y, \lambda, \theta, \psi, \sigma, \gamma)$ is derived, where λ represents wavelength and θ represents the orientation. In this example, we consider a combined, complex-number approach, not separately filtering the imaginary and real parts. This is defined in equation 1 where $x' = x\cos(\theta) + y\sin(\theta)$ and $y' = x\sin(\theta) + y\cos(\theta)$.

$$g(x, y, \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\exp\left(i\left(2\pi\frac{x'}{\lambda}+\psi\right)\right) \quad (1)$$

We use parameters $\theta \in [0, 45, 90, 135], \sigma = 5, \psi = 90, \lambda = 50$ and a kernel size of 21. These values were decided upon through testing the impulse response given by Equation 1 against viable frames with different parameters until a whole-texture pattern was obtained.

Sub-band histograms taken of the impulse responses at each rotation are used as a feature vector of the target training frame. Using the four responses together, Local Binary Patterns (LBP) [1, 4] are used to model local texture information on the final texture representation. Three sets of values for radius r and points p are evaluated: $(r = 1, p = 8)$, $(r = 2, p = 12)$ & $(r = 3, p = 16)$.

Support Vector Machines (SVM) are binary classifiers, but can be used to classify multiple classes; in order to achieve this, a cluster of SVMs are trained to form a SVC (Support



(a) Original Frame (b) $\theta = 0$ (c) $\theta = 45$

(d) $\theta = 90$ (e) $\theta = 135$

Fig. 4. Impulse responses based on input frame, the four rotation values of θ used, and the resultant output.

Vector Classifier). These are trained using both linear (SVCLIN) and radial basis function (SVCRBF) kernels, using our selected features. This results in a cluster of SVMs in a hierarchical formation. These are compared with kNN classifiers using the ball tree algorithm with $k = 5$ [10].

A testing & training strategy of 10:90 n-fold cross validation was used, where 10% of a given class' frames from within a video were selected at random and used as its training data. Figure 5 illustrates the datasets, experiment setup, and the video subdivision into the following categories: *training*, *testing* and *other*. *Other* is for any frames of class 0 which have been recorded as not containing relevant information.

3. RESULTS

All preliminary experiment results are noted in Table 2, grouped by video. Reviewing these, kNN classifiers perform with greater accuracy than SVCs. The difference between a kNN approach and SVCLIN is as high as 35% increase in successful classification. Using the same features and the same video, SVCLIN outperforms SVCRBF by up to 79% success rate. These findings demonstrate that the problem responds well to a clearly linear and isolated classification system, which remains constant irrespective of the feature being trained.

Features which are dependent on texture result in successful classification on par, or below, the statistical frame histogram metrics. LBPs do not follow a clear correlation of effectiveness based on the radius and number of points used, and sub-band histograms perform better than LBP where comparable data is present.

4. CONCLUSION

We have evaluated several machine learning and computer vision techniques as preliminary steps in understanding underwater environments. The complexity of our classification schema is sufficient for the identification of seabed substrate types, however more detailed analysis as per the CCW schema requires further additional non-visual infor-

Table 3. Experiment Results: Preliminary results of research. All values are percent of testing frames correctly classified as per our classification schema (where 0 indicates no data and B. refers to the Bangor dataset).

Video	B. Site 2			B. Site 3			B. Site 10			FoCB 1			FoCB 2					
Feature & Classifier	SVCRBF	SVCLIN	KNN	SVCRBF	SVCLIN	KNN	SVCRBF	SVCLIN	KNN	SVCRBF	SVCLIN	KNN	SVCRBF	SVCLIN	KNN	Min	Median	Max
Histogram (Greyscale)	0	0	0	0	0	97	0	0	86	53	93	97	54	71	83	53	85	97
Histogram (Colour)	0	0	98	0	0	0	0	0	88	53	96	98	54	79	86	53	87	98
Gabor Sub-band Histogram	0	0	96	0	0	0	0	0	79	0	85	99	0	68	91	68	88	99
Gabor LBP ($r = 1, p = 8$)	0	0	95	0	0	0	0	0	74	53	72	99	54	61	69	53	71	99
Gabor LBP ($r = 2, p = 12$)	0	0	95	100	0	0	0	0	75	0	0	91	0	0	0	75	93	100
Gabor LBP ($r = 3, p = 16$)	0	0	0	100	0	0	0	0	76	53	85	93	0	0	0	53	85	100
Min	0	0	95	100	0	97	0	0	74	53	72	91	54	61	69			
Median	0	0	97	100	0	97	0	0	79	53	0	98	54	70	85			
Max	0	0	98	100	0	97	0	0	88	53	96	99	54	79	91			

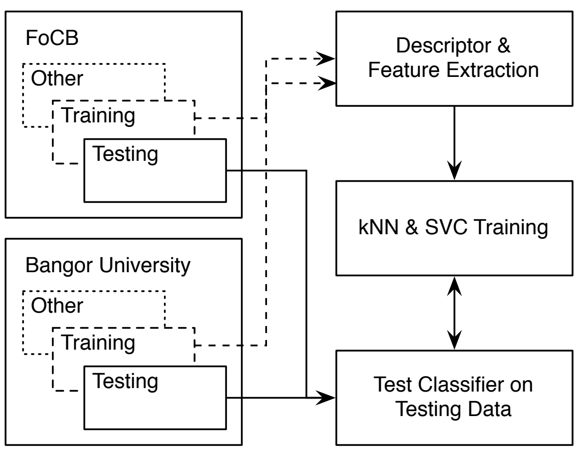


Fig. 5. Training & testing methodology for both datasets

mation. Approximating this information could potentially be performed by identifying regions for further analysis, and observing specific fauna.

We have shown that it is possible to use existing methods to achieve suitable classifiers on underwater video, with a number of caveats. The results gathered indicate a higher than anticipated rate of success using full-frame classification methods, in the case of using an SVCRBF classifier and LBP patterns with $r = 3, p = 16$, 100% testing success is noted. A key reason for this is that the videos selected from each dataset were those deemed to be most complex by a marine biologist. Complex being defined as most changes in the observed sea-bed substrates. However, even in these chosen videos the range of visually-distinct substrates as per our schema is low. This is due to the short length of the videos themselves and their coverage not being in areas pre-selected for likelihood of substrate complexity. The results validate the use of these methods to build a more generic underwater substrate classification system but do not yet give any indication as to what its accuracy could be. Given the acknowledged shortcomings, the results presented in this paper should be viewed as exploratory findings only.

The use of different image descriptors and evaluation of different texture extraction methods (and parameters thereof) could provide more successful classification. In particular, shape and approximation of motion against a static background could theoretically eliminate issues in existing video relating to trawler-impact debris via automated dismissal. The most prominent direction for continued research is the automatic isolation of regions-of-interest (suitable illumination, accurately masking any visible recording equipment present in frames) rather than full-frame processing.

Furthermore, the project aims to investigate the use of ROVs in obtaining footage for further analysis. The collection of meta-data including depth, water temperature and GPS coordinates during ROV survey will enable research into more detailed mapping of the sub-aquatic environment. Sea-bed depth and temperatures are also key factors in identifying habitats for species of interest to marine biologists which, in conjunction with the work presented in this paper, could lead to accurate, automated systems for surveying in the future.

Acknowledgments

This research is gratefully funded by KESS, which is part-funded by the European Social Fund (ESF) through the European Union's Convergence Programme (West Wales and the Valleys) administered by the Welsh Government.

We offer our thanks to the FoCB for their commitment to this research and continued input. We also wish to thank Dr Jan Hiddink and his research team at Bangor University for kindly allowing us use of their video collection.

5. REFERENCES

[1] Ma Shiela Angeli Marcos, Laura David, Eileen Peñaflor, Victor Ticzon, and Maricor Soriano, "Automated benthic counting of living and non-living components in ngedarrak reef, palau via subsurface underwater video," *Environmental monitoring and assessment*, vol. 145, no. 1-3, pp. 177–184, 2008.

[2] JG Hiddink, S Jennings, MJ Kaiser, AM Queirós, DE Duplisea, and GJ Piet, "Cumulative impacts of seabed trawl disturbance on benthic biomass, production, and species richness in different habitats," *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 63, no. 4, pp. 721–736, 2006.

[3] Bernhard Riegl, Jan L Korrubel, and Charles Martin, "Mapping and monitoring of coral communities and their spatial patterns using a surface-based video method from a vessel," *BULLETIN OF MARINE SCIENCE-MIAMI-*, vol. 69, no. 2, pp. 869–880, 2001.

[4] Maricor Soriano, Sheila Marcos, Caesar Saloma, Miledel Quibilan, and Porfirio Alino, "Image classification of coral reef components from underwater color video," in *Oceans, 2001. MTS/IEEE Conference and Exhibition*. IEEE, 2001, vol. 2, pp. 1008–1013.

[5] Dirk Walther, Duane R Edgington, and Christof Koch, "Detection and tracking of objects in underwater video," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. IEEE, 2004, vol. 1, pp. I–544.

[6] Gayathri Nadarajan, Cheng-Lin Yang, and Yun-Heh Chen-Burger, "Multiple ontologies enhanced with performance capabilities to define interacting domains within a workflow framework for analysing large undersea videos," in *International Conference on Knowledge Engineering and Ontology Development*, 2013.

[7] Concetto Spampinato, Simone Palazzo, Daniela Giordano, Isaak Kavasidis, Fang-Pang Lin, and Yun-Te Lin, "Covariance based fish tracking in real-life underwater environment.," .

[8] Countryside Council for Wales, "Groups of marine habitats found around the coasts and sea of wales for use in assessing sensitivity to different fishing activities," Tech. Rep., Welsh Government, 2010.

[9] V Shiv Naga Prasad and Justin Domke, "Gabor filter visualization," *J. Atmos. Sci*, vol. 13, 2005.

[10] Nitin Bhatia et al., "Survey of nearest neighbor techniques," *arXiv preprint arXiv:1007.0085*, 2010.