Smart Multi-Modal Marine Monitoring via Visual Analysis and Data Fusion

Dian Zhang
Clarity, Dublin City University, Dublin9, Ireland
dian.zhang2@mail.dcu.ie

Edel O'Connor
Marine Institute, Galway, Ireland
edel.oconnor-ext@marine.ie

Timothy Sullivan
Marine and Environmental Sensing Technology Hub, Dublin City University, Dublin 9
tim.sullivan@dcu.ie

Kevin McGuinness
Clarity, Dublin City University, Dublin9, Ireland
kevin.mcguinness@dcu.ie

Fiona Regan
Marine and Environmental Sensing Technology Hub, Dublin City University, Dublin 9
fiona.regan@dcu.ie

Noel E. O'Connor
Clarity, Dublin City University, Dublin9, Ireland
noel.oconnor@dcu.ie

ABSTRACT

Estuaries and coastal areas contain increasingly exploited resources that need to be monitored, managed and protected efficiently and effectively. This requires access to reliable and timely data and management decisions must be based on analysis of collected data to avoid or limit negative impacts. Visually supported multi-modal sensing and data fusion offer attractive possibilities for such arduous tasks. In this paper, we demonstrate how an *in-situ* sensor network can be enhanced with the use of contextual image data. We assimilate and alter a state-of-the-art background modelling technique from the image processing domain in order to detect turbidity spikes in water quality sensor measurements automatically. We then combine this with visual sensing to identify abnormal events that are not caused by local activities. The system can potentially assist those charged with monitoring large scale ecosystems, combining real-time analytics with improved efficiency and effectiveness.

Categories and Subject Descriptors

I.4 [**IMAGE PROCESSING AND COMPUTER VISION**]: Miscellaneous; I.5.4 [**Applications**]: Computer vision

Keywords

data fusion, environmental monitoring, multi-modal sensing, spike detection, visual sensing

1. INTRODUCTION

Increasing demands from various scientific and management communities for monitoring issues such as ecosystem change, climate change, water quality, coastal erosion and flooding along with legislative requirements at both national and international level has led to a need for innovative research into large-scale, reliable and sustainable monitoring of marine and freshwater environments. In particular, the need to continuously protect and regulate these environments is being recognised by governments and policy makers.

On a more regional scale, coastal and freshwater environments represent vital assets on many levels and need continuous monitoring and protection. Modern sensing technologies, such as autonomous wireless sensor networks (WSN), provide an opportunity to meet the challenges of high spacial and temporal scales. However, it is becoming increasingly clear that in order to adequately monitor marine and freshwater environments, they need to be characterised from multiple perspectives. Environmental events can be only detected and modelled accurately from multiple diverse data sources. Only through combining the perspectives offered by a variety of multi-modal information sources can we obtain a true picture of events. Previous work have shown that video systems are effective tools for coastal environment monitoring [3, 8]. Other studies have investigated the use of cameras in other ecological monitoring applications. In [9], Richardson et al. investigate the use of cameras to monitor spring green-up in a deciduous northern hardwood forest. Graham et al. [4] explored whether cameras can be used to determine the dynamics of expanding leaf area for *Rhododendron occidental*. They concluded that cameras provide an inexpensive tool to quantify ecological dynamics.

In our previous work [14], we have shown that shipping activities in ports greatly affects the aquatic ecosystem. We successfully identified the presence of a ship within an image in a challenge dataset. However, the effects of these events could differ depending on the different motion patterns of ships when entering or leaving the harbour. The principle limitation of this previous work was an inability to differentiate moving vessels from static vessels. In this work, our focus is on creating a smarter marine environmental monitoring network by fusion of visual sensor modality and *in-situ* wireless sensors. A more sophisticated space-time feature, dense trajectories, is applied to classify the motion of a shipping event. We also automate the detection of turbidity spikes from *in-situ* sensor modality and then combine with

MAED'13, October 21, 2013, Barcelona, Spain.

ACM 978-1-4503-2401-4/13/10 ...$15.00.
http://dx.doi.org/10.1145/2509896.2509903.

shipping trajectory detection results from visual sensors to discover the true abnormal events requiring further analysis.

2. SITE OVERVIEW

Poolbeg Marina (latitude: 53°20'39", longitude: -6°12'59") is located on the lower Liffey Estuary in Dublin Ireland (Figure 1). The estuary hosts a diverse ecosystem including benthic communities, fish and shellfish, seabird populations and marine mammals [10, 13]. The topography of the estuary has been greatly modified, and is constrained by walls along its whole length and is regularly dredged to remove accumulated sediments. The working site is located in the upper part of the Estuary, where the ship traffic is less intensive. Average water depth in the area is approximately $8m$ and the width of the channel is approximately $260m$. Due to the large amount of activity at the site and its importance from an environmental and ecological perspective, the site was equipped with a multi-parameter *in-situ* sensor along with a visual sensing system. To better understand the ecosystem at the site, we first focus on the turbidity values as it is the most complicated water quality parameter compared to others. It can be affected by many factors such as rainfall, wind, tides and shipping traffic. Turbidity is defined as the decrease in the transparency of a solution owing to the presence of suspended and some dissolved substances, which causes incident light to be scattered, reflected and attenuated rather than transmitted in straight lines [1]. Turbidity is now seen as a key water pollutant and often used as a surrogate variable for suspended solids concentration [2]. The negative impacts of increased turbidity are well defined in the literature [7].

Shipping can regularly and profoundly affect turbidity levels through a number of mechanisms, including shore erosion from wakes, increased vertical mixing and stirring of the sediments, especially in the turning area outside harbours from large ships (150-200m long) or indirectly through regular dredging [12]. Pressure changes, propeller suction, use of bow thrusters, drag and acceleration caused by shipping can all result in visible water displacement, swell, pressure waves and turbulence. All of this results in periodic increases in mixing energy driving vertical mixing, artificial upwelling, temporary water currents and material transport [6].

Figure 1: Poolbeg Marina overview with the location of deployed camera and *in-situ* sensor – from Google maps

A YSI 6600EDS sonde equipped with turbidity sensor was deployed at a depth of $2.5m$ from the water surface. The sonde was validated using a portable turbidity meter Turb® 430 IR (VWR Ireland) which was calibrated in the laboratory. Sensor maintenance was undertaken weekly and copper tape and mechanical wipers were used to control biofouling on the probe. Along with the *in-situ* sonde, an IP-66 rated Axis P1344-E network camera was also deployed at the site. The camera was mounted on a pole at a height of $4.36m$ above the ground and approximately $20m$ from bank wall. The visual sensor continuously sends images to a cloud server through 3G mobile broadband at approximately 1 frame every 10 seconds. However, this frame work is not guaranteed due to an unreliable network connection.

3. METHODOLOGY

3.1 Turbidity Spike Detection

A turbidity spike is defined as the comparatively large upward movement followed by downward movement of a single or several turbidity values in a short period. In order to detect a turbidity spike, we alter the pixel-based adaptive segmenter (PBAS) method originally proposed by Martin Hofmann for image segmentation [5]. To classify a new incoming turbidity value $I(t)$ at time t, a turbidity background model consisting of recently observed background values is built. The decision is made based on a dynamic threshold $T(t)$. If the incoming value at time t is greater than the threshold at time t, it will be classified as a spike value. If a series of turbidity samples are greater than the threshold, they will be merged into a spike event. However, a spike event can also consist of a single turbidity spike value that is outstanding from its neighbours. The background turbidity model is updated over time in order to handle gradual changes such as the effect from rainfall and tides. The update rate is controlled by a learning parameter $L(t)$. The essential and novel idea of the PBAS approach is that both the decision threshold $T(t)$ and background learning rate $L(t)$ are dynamically adapted based on an estimate of the background dynamics. Moreover, these parameters can be used as the controller of the detection system sensitivity. However, due to the nature of water turbidity values, we altered the original method to suit one dimensional sensor measurements.

The background model is defined by an array of N recently observed values.

$$B(t) = \{B_1(t), ..., B_k(t), B_N(t)\} \tag{1}$$

In [5], incoming values are classified based on the number of distances between input value $I(t)$ and all elements in $B(t)$ that are smaller than threshold $T(t)$. We found that just comparing the minimum distance with the threshold is sufficient to differentiate the measurements.

$$I(t) = \begin{cases} 1, & \text{if } \min(\ \text{dist}(I(t),\ B_k(t))\ ) > T(t) \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

If the input value is classified as background ($I(t) = 0$), it can be used for updating the model. The update probability depends on the learning rate $L(t)$.

When monitoring turbidity values in a marine environment, there can be time periods with high dynamics, such as after a heavy rainfall, and time periods with little variations. Ideally for highly dynamic time periods, the threshold $T(t)$ should be increased and for time periods with low variations, $T(t)$ should be decreased. In order to quantify the turbidity dynamics, the average of the previous N minimum distances between input turbidity value and best matching background value are kept and the average $\overline{d}_{min}(t)$ of these

values is the measure of the background variations. The decision threshold can then be adapted as follows:

$$T(t) = \begin{cases} T(t) \times (1 - T_{inc/dec}), & \text{if } T(t) > \overline{d}_{min(t)} \times T_{scale} \\ T(t) \times (1 + T_{inc/dec}), & \text{otherwise} \end{cases} \tag{3}$$

Both $T_{inc/dec}$ and T_{scale} are fixed parameters. $T_{inc/dec}$ is the update step and T_{scale} is the controller of the update rate.

Another important parameter of the method is the background learning rate $L(t)$ which is defined as follows:

$$R(t) = \begin{cases} R(t) + \frac{L_{inc}}{\overline{d}_{min}(t))}, & \text{if spike} = \text{true} \\ R(t) - \frac{L_{dec}}{\overline{d}_{min}(t))}, & \text{if spike} = \text{false} \end{cases}$$

$$L(t) = \frac{1}{R(t)/R_{upper}} \tag{4}$$

Turbidity has very unique characteristics. Turbidity values measured by *in-situ* sensors are typically very noisy and vary from a baseline (they change gradually due to "global" effects, such as rainfall, tides, season etc.). Estimating this dynamic baseline is cumbersome and the model needs to be updated frequently. Due to the above reasons, we altered the original method in [5]. Turbidity spikes always introduce large variations from the baseline. Thus, when a spike occurs, the back ground model should be updated slowly or not be updated at all. When turbidity falls back to the baseline, the model should be updated quickly. When a turbidity spike value is first detected after a low variation period ($\overline{d}_{min}(t)$ is small), $R(t)$ increases rapidly, thus the learning rate $L(t)$ decreases sharply. However, $\overline{d}_{min}(t)$ will become large quickly when multiple spike values are detected which results in $R(t)$ and indeed $L(t)$ remaining constant or changing slightly. When sediment settles, $\overline{d}_{min}(t)$ becomes small and $L(t)$ will increase. The variation of R is limited by an upper and lower bound $R_{lower} < R(t) < R_{upper}$, thus, it cannot go out of a specified range. The learning rate $L(t)$ is used as the update probability and an element in the background model is randomly selected and replaced by the incoming value. R_{inc} and R_{dec} are fixed values that separate the rising and falling rate of $R(t)$. However, the calculation of this learning rate is task dependant. For example, if the user wants to model the overall trend, the normalised $R(t)$ can be used as the update probability.

3.2 Ship Trajectory Detection

Since our objective is to be able to detect when large ships enter the harbour that may cause a turbidity spike event, we need to extract a set of features that are sufficiently discriminative to allow us to classify such events.

Firstly, a dense trajectory feature [11] is extracted from each frame. The dense trajectory feature describes the motion pattern of an interesting point in an image. However, each image may produce varying numbers of such descriptors. In order to compare these features and further classify shipping events, a vector quantization method known as bag-of-visual-words is adopted. Each dense feature is passed to a K-means clustering method and represented by the clustering centre ("a visual word") to which it belongs. The normalized histogram of all the visual words within an N minute time window is calculated as the feature input of the classifier. A ship-entering event may occur partially in a time window. To avoid this issue, time windows overlap each other by a small fraction of the window length. Thus, an event that is not fully covered by previous time windows may be caught by the next time phase. Classification is performed by Support Vector Machine with Radial Basis Function (RBF) kernel.

4. EXPERIMENTS AND RESULTS

4.1 Turbidity Spike

The turbidity data used for the following experiments were measured by the multi-parameter sonde between 28th Oct, 2012 and 25th Nov, 2012 with sampling rate of $15mins$. In order to synchronize the output with results from ship trajectory detection, only a sub-section of the results (7 days from 19th Nov and 25th Nov 2012) are examined. The following parameter values were used: $N = 15$, $L_{lower} = 0.1$, $L_{upper} = 3.0$, $L_{inc} = 1.0$, $L_{inc} = 0.1$, $R_{inc/dec} = 0.05$, $R_{scale} = 4.5$. Figure 2 shows a 7 day sub-section of turbidity spike detection results. The red dots are the turbidity spikes detected, blue dashed line is the sensor measurements and the green solid line is the closest matching entry in the background model. As illustrated in Figure 2, most of the abnormal turbidity readings are detected accurately. Figure 3 demonstrates the adaptation of classification threshold and background learning rate based on the variation of the average minimum distance between input value and background model. The blue dashed line is the decision threshold. The minimum distance between real value and closest match background entry is shown in cyan and the red solid line at the bottom is the background learning rate. The decision threshold and background learning rate are adapted based on the trend of the input values. When the turbidity is noisy, the threshold rises quickly. However, when turbidity is stable, the threshold decreases slowly as expected. In contrast, the background model learning rate decreases rapidly when a spike occurs and increases slowly when no spike arises.

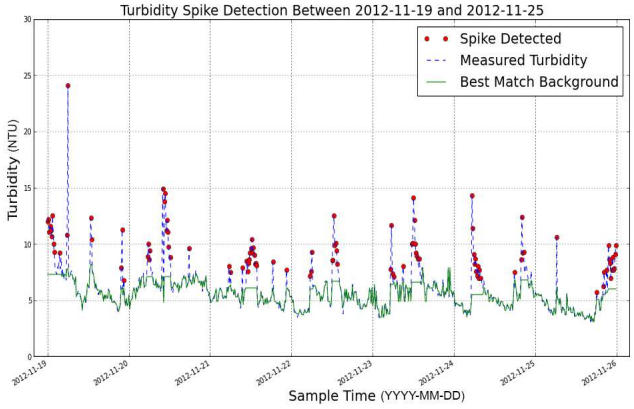

Figure 2: A 7 day sub-section of turbidity spike detection results

4.2 Shipping Entering Events

22 days of image data from 28th Oct, 2012 to 18th Nov, 2012 were used to build a model and 7 days of image data between 19th Nov, 2012 and 25th Nov, 2012 were used for

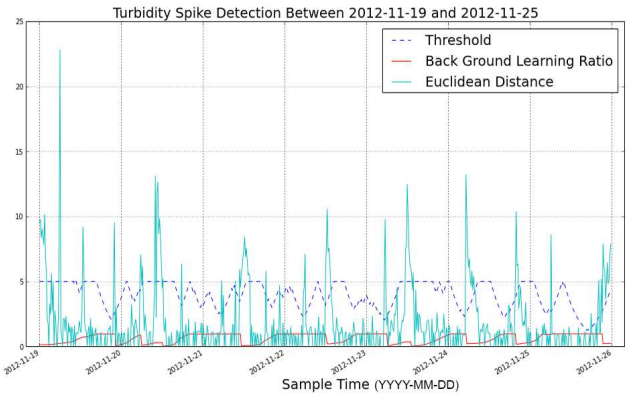

Figure 3: Classification threshold, background model learning rate and Euclidean distance between input value and best match background element

evaluation. The data exhibited a wide variety of lighting and weather conditions, as well as a variety of ship trajectories. Figure 4 demonstrates the complexity of the dataset.

Figure 4: Sample image data

Over $255,000$ color images of 640×480 pixels were annotated as the ground truth of the dataset. These images were grouped into $15mins$ time intervals with $5mins$ overlapping (e.g 2012-Nov-10 15:00 to 2012-Nov-10 15:15 and 2012-Nov-10 15:05 to 2012-Nov-10 15:20). If a large part of the event falls into a time interval, it will be annotated as true. Thus, the same shipping event may lead to multiple positive entries in the dataset. However, these entries are temporally close to each other. The total amount of shipping entering events in the training and testing set are 54 and 17 respectively, the total positive samples are 134 and 44. To reduce the amount of data that needs to be processed, a region of interest $(x : 0, y : 100, width : 640, height : 200)$ is drawn on the original image before extracting features. The number of visual words is set to 100. Classification is performed in the WEKA data analysis environment using LIBSVM implementation with default parameters. To avoid over fitting, 10-fold cross-validation was used when building the model.

no ship entering	ship entering	classified as
1968	4	no ship entering = 0
13	31	ship entering = 1

Table 1: Confusion Matrix of Testing Data

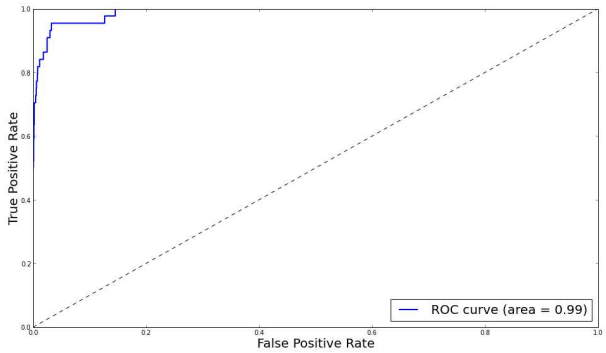

Figure 5: ROC curve of ship traffic classification

The classification confusion matrix is shown in Table 1 and the receiver operating characteristic (ROC) curve is shown in Figure 5. As can be seen, 4 out of 1972 negative samples and 13 out of 44 positive samples are misclassified. As previously discussed, a single ship entry event may result in multiple true samples in the dataset. For the purpose of this work, if any data entry within the event time period is classified as true, we can safely assume the event is detected. For example, if a ship started entering the port at 15:03 and docked at 15:12, both data entry with timestamps 15:00 to 15:15 and 15:05 to 15:20 in the dataset would be annotated as true but if either of these two data entries are classified correctly we assume the event is detected. It is also acceptable if a false positive sample is next to a true positive sample. This results in 1 missed event out of 17 ship entries and 1 event (1 false positive entry is next to a true positive sample, the other 3 false positive entries are temporally next to each other and grouped as an event) incorrectly classified. Figure 6 shows some example images of incorrectly classified events. As can be seen, the false positive sample was caused by a medium-sized cargo ship entering on the opposite side of the port, which is not of interest. Additionally, the missed event occurs in the dark $(4am)$, which may have been caused by insufficient production of interesting points for trajectory classification.

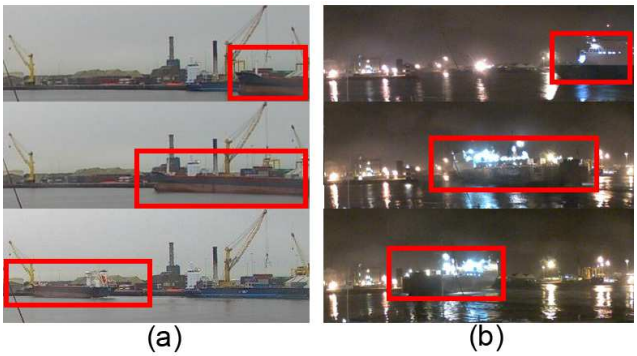

Figure 6: Sample images of (a) incorrectly classified and (b) missed shipping events

4.3 Data Fusion

To better understand and subsequently model the environmental dynamics at the observation site, the output of vi-

sual sensor data was combined with the spike events detected to further distinguish whether a spike event is caused by shipping traffic. This filters out the turbidity spikes caused by local activities and determines the real abnormal events that require further analysis. The turbidity spike points detected (as shown in Figure 2) are manually grouped into events by timestamps. This results in 20 events (the partial event at the beginning is not included). Figure 7 shows turbidity events in combination with shipping events. From

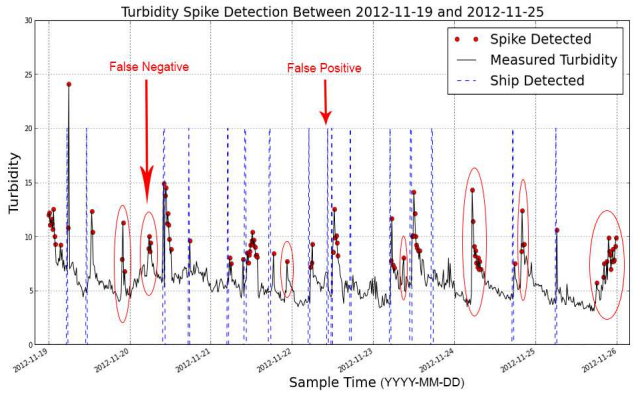

Figure 7: Matching of shipping event and turbidity spikes by their timestamps

the graph, it can be seen that 13 out of 20 turbidity events are very likely caused by shipping traffic, which results in only a small set of abnormal events requiring further analysis. A single turbidity event caused by shipping traffic is not detected due to misclassification, however, the operator could simply examine image data manually and remove this event. Thus, only 6 (30%) events out of 20 need to be further studied.

5. CONCLUSIONS

In this work, we used state-of-the-art action recognition methods to detect shipping which may cause rapid changes in water turbidity levels. We also adapted a background modelling technique from the image processing domain to detect sudden changes in 1-D water quality sensor data. By fusion of multiple sensor modalities into an integrated smart system, we have demonstrated the potential of this approach in assisting in better understanding and subsequent modelling of observation sites more efficiently and effectively. The visual sensing data has provided augmentative intelligence to help understand the causes of some large spikes in turbidity values. Turbidity occurrences can then be further classified as "caused by local activities" or "cause unknown" which results in only the remaining true abnormal events requiring further analysis.

6. ACKNOWLEDGEMENTS

QUESTOR, MESTECH, CLARITY

7. REFERENCES

[1] S. G. J. Aarninkhof, I. L. Turner, T. D. T. Dronkers, M. Caljouw, and L. Nipius. Nearshore subtidal bathymetry from time-exposure video images. *Geophysical Research*, 110 (C6), 2005.

[2] H. Chanson, M. Takeuchi, and M. Trevethan. Using turbidity and acoustic backscatter intensity as surrogate measures of suspended sediment concentration in a small subtropical estuary. *Journal of Environmental Management*, 88(4):1406 – 1416, 2008.

[3] A. F. S. Edel O'Connor. *Trust and reputation in multi-modal sensor networks for marine environmental monitoring*. PhD thesis, Dublin City University, March 2012.

[4] E. A. Graham, E. M. Yuen, G. F. Robertson, W. J. Kaiser, M. P. Hamilton, and P. W. Rundel. Budburst and leaf area expansion measured with a novel mobile camera system and simple color thresholding. *Environmental and Experimental Botany*, 65(2-3):238 – 244, 2009.

[5] M. Hofmann, P. Tiefenbacher, and G. Rigoll. Background segmentation with feedback: The pixel-based adaptive segmenter. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 38–43, 2012.

[6] Lindholm, Tore and Svartström, Mia and Spoof, Lisa and Meriluoto, Jussi. Effects of ship traffic on archipelago waters off the långnäs harbour in Åland, sw finland. *Hydrobiologia*, 444(1-3):217–225, 2001.

[7] C. Newcombe and D. MacDonald. Effects of suspended sediments on aquatic ecosystems. *North American Journal of Fisheries Management*, 11(1):72–82, 1991.

[8] E. O'Connor, A. Smeaton, and N. O'Connor. A multi-modal event detection system for river and coastal marine monitoring applications. In *OCEANS, 2011 IEEE - Spain*, pages 1 –10, june 2011.

[9] A. Richardson, J. Jenkins, B. Braswell, D. Hollinger, S. Ollinger, , and M. Smith. Use of digital webcam images to track spring green-up in a deciduous broadleaf forest. *Ecosystem Ecology*, 152(2):323–334, 2007.

[10] S. Roth and J. G. Wilson. Functional analysis by trophic guilds of macrobenthic community structure in dublin bay, ireland. *Journal of Experimental Marine Biology and Ecology*, 222(1?2):195 – 217, 1998.

[11] H. Wang, A. Klaser, C. Schmid, and C.-L. Liu. Action recognition by dense trajectories. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 3169–3176, 2011.

[12] D. H. Wilber and D. G. Clarke. Biological effects of suspended sediments: A review of suspended sediment impacts on fish and shellfish with relation to dredging activities in estuaries. *North American Journal of Fisheries Management*, 21(4):855–875, 2001.

[13] J. Wilson. Productivity, fisheries and aquaculture in temperate estuaries. *Estuarine, Coastal and Shelf Science*, 55(6):953 – 967, 2002.

[14] D. Zhang, E. O'Connor, K. McGuinness, N. E. O'Connor, F. Regan, and A. Smeaton. A visual sensing platform for creating a smarter multi-modal marine monitoring network. In *Proceedings of the 1st ACM international workshop on Multimedia analysis for ecological data*, MAED '12, pages 53–56, New York, NY, USA, 2012. ACM.