Methods in Ecology and Evolution

Methods in Ecology and Evolution 2017, **8,** 996–1004 doi: 10.1111/2041-210X.12746

Expert, Crowd, Students or Algorithm: who holds the key to deep-sea imagery 'big data' processing?

Marjolaine Matabos*,[1], Maia Hoeberechts[2,3], Carol Doya[4], Jacopo Aguzzi[4], Jessica Nephin[2], Thomas E. Reimchen[5], Steve Leaver[5], Roswitha M. Marx[5], Alexandra Branzan Albu[3,6], Ryan Fier[6], Ulla Fernandez-Arcaya[4] and S. Kim Juniper[2,7]

[1]*Ifremer, Centre de Bretagne, REM/EEP, Laboratoire Environnement Profond, 29280 Plouzané, France;* [2]*Ocean Networks Canada, University of Victoria, Victoria, BC V8W 2Y2, Canada;* [3]*Department of Computer Science, University of Victoria, Victoria, BC, Canada;* [4]*Institute of Marine Sciences (ICM-CSIC), Paseo Marítimo de la Barceloneta 37-49, Barcelona 08003, Spain;* [5]*Department of Biology, University of Victoria, Victoria, BC, Canada;* [6]*Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, Canada; and* [7]*School of Earth and Ocean Sciences and Biology Department, University of Victoria, Victoria, BC, Canada*

Summary

1. Recent technological development has increased our capacity to study the deep sea and the marine benthic realm, particularly with the development of multidisciplinary seafloor observatories. Since 2006, Ocean Networks Canada cabled observatories, have acquired nearly 65 TB and over 90 000 h of video data from seafloor cameras and remotely operated vehicles. Manual processing of these data is time-consuming and highly labour-intensive, and cannot be comprehensively undertaken by individual researchers. These videos are a crucial source of information for assessing natural variability and ecosystem responses to increasing human activity in the deep sea.

2. We compared the performance of three groups of humans and one computer vision algorithm in counting individuals of the commercially important sablefish (or black cod) *Anoplopoma fimbria*, in recorded video from a cabled camera platform at 900 m depth in a submarine canyon in the Northeast Pacific. The first group of human observers were untrained volunteers recruited *via* a crowdsourcing platform and the second were experienced university students, who performed the task for their ichthyology class. Results were validated against counts obtained from a scientific expert.

3. All groups produced relatively accurate results in comparison to the expert and all succeeded in detecting patterns and periodicities in fish abundance data. Trained volunteers displayed the highest accuracy and the algorithm the lowest.

4. As seafloor observatories increase in number around the world, this study demonstrates the value of a hybrid combination of crowdsourcing and computer vision techniques as a tool to help process large volumes of imagery to support basic research and environmental monitoring. Reciprocally, by engaging large numbers of online participants in deep-sea research, this approach can contribute significantly to ocean literacy and informed citizen input to policy development.

Key-words: computer vision algorithms, crowdsourcing, deep-sea imagery, Digital Fishers, fish counting, Ocean Networks Canada, seafloor observatories, underwater video

Introduction

Advances in instrumentation are allowing ecosystems to be investigated at increasing spatial and temporal resolution (Porter *et al*. 2009). As a direct result, researchers in the environmental and biological sciences are faced with growing challenges and opportunities related to 'big data' (Grémillet *et al*. 2012; Woodward *et al*. 2014). Data are accumulating faster than the processing power of research laboratories and institutions, and their effective exploitation requires more human resources and additional computational solutions. Computer algorithms have proven to be effective at assimilating and summarizing large volumes of scalar data (e.g. Belkin & O'Reilly 2009), but computer vision software solutions are still far from replacing the human eye in extracting scientific information from complex data types like imagery (Aguzzi *et al*. 2009; Purser *et al*. 2009; Aron *et al*. 2010; Schoening *et al*. 2012). For some image analysis applications, engaging the public in initial data processing or annotation (i.e. adding caption and metadata to a digital image) has yielded useful results. The astronomical science community was among the first to apply crowdsourcing approaches to image analysis, engaging the public in analysing a huge

*Correspondence author. E-mail: marjolaine.matabos@ifremer.fr

archive of space imagery through the Zooniverse platform (https://www.zooniverse.org/projects, Galaxy Zoo, Lintott *et al.* 2008). Crowdsourcing has become a form of citizen science where members of the public contribute to scientific research projects by acquiring and/or processing data, with few prerequisite knowledge requirements (Silvertown 2009). Crowdsourcing has benefited from the Web 2.0 technologies that enabled user-generated content and interactivity, such as wiki pages, web apps or social media. These web developments have enabled structured data analysis by a substantial number of online contributors (Wiggins & Crowston 2011).

Crowdsourcing has the potential to contribute to biological studies that use deep-sea video and still photo imagery as a primary source of information. The floor of the deep ocean, and its important but still unquantified reservoir of biodiversity, are invisible from space and can only be imaged from a few metres distance using artificial lighting and deep-sea cameras. As a result, only about 5% of the seabed has been surveyed by platforms like remotely operated vehicles (ROVs) and autonomous underwater vehicles (AUVs) (Ramirez-Llodra *et al.* 2010). *In situ* imagery allows biologists to quantify the spatial distribution and seasonal variability of deep-sea species in their natural habitat, and to document their behaviour (Tunnicliffe 1990; Copley *et al.* 1997, Copley, Jorgensen & Sohn 2007; Aguzzi *et al.* 2010; Porteiro *et al.* 2013). Seafloor observatories currently under development or in operation in several areas of the world ocean will produce unprecedented volumes of imagery that will create a processing bottleneck. The NEPTUNE and VENUS cabled observatories, operated by Ocean Networks Canada (ONC; http://oceannetworks.ca) off Vancouver Island, Canada, support continuous observations of faunal and habitat variables and have been recording daily video imagery from coastal to abyssal habitats since February 2006. The rapidly growing data archive now contains video from 26 current and historical video camera systems across the network, whose output, when added to ROV imagery from observatory installation and maintenance operations, currently consists of over 90 000 h of video for a total of nearly 65 TB of video data.

The field of computer vision is well-developed for certain land-based image analysis tasks such as, among others, human facial recognition (Zafeiriou, Zhang & Zhang 2015) and human behaviour analysis (Vishwakarma & Agrawal 2012). In contrast, underwater imagery analysis is an emerging field that presents unique challenges not found in other domains, such as light propagation effects in water (i.e. differential spectral attenuation, scattering) and non-uniform artificial lighting, to name a few (Schettini & Corchs 2010). Most automated techniques are designed to sort images based on predetermined criteria or to annotate images to add information about objects or areas of interest. They vary from semi-automatic methods, which require various degrees of human intervention during execution, to automatic methods which, once algorithms are trained using manually generated training sets, can sort or produce annotations without human intervention (e.g. Chuang, Hwang & Williams 2014). Best analytical results are achieved when automated techniques are developed for each specific target application and dataset, as these techniques often do not generalize easily.

Deep-sea citizen science is still in its infancy, and it is difficult to evaluate its potential for contributing to our knowledge of this environment. Only two crowdsourcing applications for underwater seafloor imagery are widely available to date (i.e. the Zooniverse Seafloor Explorer, https://www.seafloorexplorer.org and Ocean Networks Canada's Digital Fishers, http://dmas.uvic.ca/DigitalFishers), and marine citizen science projects are relatively few compared with projects developed on land (Roy *et al.* 2012). The goal of this study was to evaluate the accuracy of crowdsourcing in relation to computer vision algorithms and human experts, in the processing of deep-sea video imagery for deep-sea biologists. We focused on identifying and counting a commercially important fish species (the sablefish *Anoplopoma fimbria*; Kulka & Pitcher 2001). A selected video dataset was screened by untrained citizen scientists, a computer vision algorithm for fish counting (Fier, Branzan Albu & Hoeberechts 2014), undergraduate university students (3rd year biology class) and a scientific expert (PhD student). Ultimately, we aim to provide guidance to researchers for optimizing the processing of imagery 'big data' in the context of a growing global network of deep-sea observatories.

Materials and methods

SAMPLING SITE AND DATA ACQUISITION

The videos analysed in this study were acquired by a camera platform (Mid-East) at a 900 m depth seabed site in Barkley Canyon, a submarine canyon in the Northeast Pacific Ocean, off Vancouver Island, Canada. For this study, 50 s of video (MP4 format) was acquired every 30 min over a 1-month period, from 21.30 h on 14 October to 00.00 h on 14 November 2011, Pacific Standard Time (PST, local time), for a total of 1439 video sequences (see Video S1, Supporting Information, for an example). The camera orientation was fixed at 45° down from horizontal, so that the field of view imaged approximately 2 m^2 of the sediment-covered seabed. The task for all human and machine participants was to count sablefish, *A. fimbria* (Fig. 1), in each video clip in the project dataset. The target species (sablefish) was easily identifiable by untrained observers, and images had few non-target fish species. This dataset formed part of a PhD study by C. Doya (Doya *et al.* 2014), referred to hereafter as the 'Expert'. For each video segment, the Expert manually reported in a spreadsheet the number of individuals of the most abundant and discernible species over the entire video, using QuickTime© media player software (Apple Inc., Cupertino, CA, USA). When a sablefish was not fully included in the region of interest or partially hidden by another fish, but was still identifiable, the animal was counted. When several sablefish overlapped and to avoid miscounting, orientation and trajectory were used to identify individuals.

UNIVERSITY STUDENT PARTICIPATION

The project dataset was provided to a class of 60 3rd year biology students as a laboratory exercise for Biology 335 (Ichthyology), at the University of Victoria in 2012. Each video clip was reviewed by 1–4 different groups of students (working in pairs). Students were asked to count individuals and identify fish species in the videos and also record

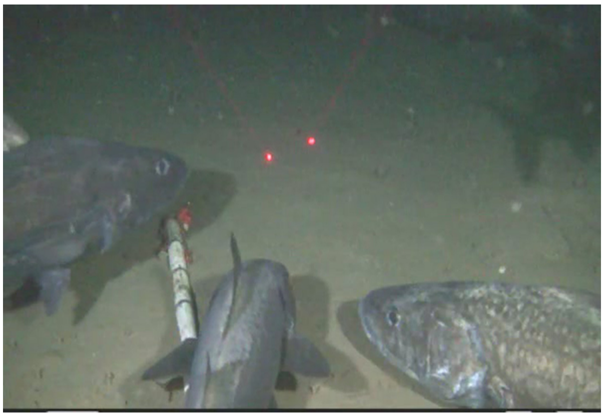

Fig. 1. Photo extracted from a video recorded in Barkley canyon, off Vancouver Island (BC, Canada) showing sablefish, *Anoplopoma fimbria.*

data on the laterality of fish behavioural response (left or right turning) to the camera structure as part of the laboratory exercise requirements (results not shown). The students involved had no background in image analysis. They were given a 10-min introduction to ocean observatories and camera systems, followed by a 15 min demo of the online data access and annotation tools. The students were then instructed on the tasks to be accomplished and the methodology, including how to recognize the species of interest. The videos were watched independently by each group of Students on their own computers. They were given a period of a few weeks to complete the tasks, outside of lecture/laboratory time. Students performed all annotations online using the ONC online annotation tool available in the video viewer SeaTube (dmas.uvic.ca/SeaTube, Fig. S1). After watching the full segment of video, students were asked to add an annotation using the dedicated button on the interface (Fig. S1). All annotations were recorded in the ONC database. Results from a student who did not annotate a single fish in all processed videos were disregarded.

CROWDSOURCING

In collaboration with the Centre for Global Studies at the University of Victoria, ONC developed *Digital Fishers* (http://dmas.uvic.ca/DigitalFishers; Hoeberechts *et al.* 2015) in 2011, an online crowdsourcing platform to help analyse and annotate video acquired from deep-sea cameras. A special 'sablefish mission' to annotate the project video dataset was conducted from May 2014 to February 2015. When connecting to the Digital Fishers platform, participants were informed though a pop-up window of the ongoing task which consisted of determining, after watching the 1-min video, how many sablefish were present. An *'ad hoc* tailored' tutorial provided cues for recognizing the species of interest, mainly through pictures. At the end of each video clip, observers were prompted to enter an observed sablefish count, which when completed allowed them to view the next clip (see Fig. S2). Clips were provided in random temporal order to the users. A button with choices from 0 to 12+ (i.e. maximum number of fish observed by the Expert) simplified the annotation task and linked participant information to counts in the database.

COMPUTER VISION ALGORITHM

A custom computer vision algorithm was developed over the course of 4 months as a computing science student project to specifically detect and count sablefish in video from the Barkley Canyon camera site (referenced as the 'Algorithm' in this paper). An overview of the method is presented here (see Fig. S3); for details, the reader is referred to Fier, Branzan Albu & Hoeberechts (2014). The approach consisted of three sequential modules: 'Preprocessing', 'Detection' and 'Tracking and Counting'. The first module (Preprocessing) used sequential application of filters, colour restoration techniques and lighting and contrast adjustments to enhance fish-related features while reducing noise in the videos. The underwater video used for this work presented challenges for automated analysis, including limited visible range, low contrast, non-uniform lighting, wavelength dependent colour attenuation, compression artifacts, light scattering by marine snow or resuspended sediment, and turbidity. The preprocessing step attempted to mitigate these effects to enhance the performance of the subsequent steps.

The second module (Detection) identified potential fish candidate regions using three separate background subtraction techniques which were combined using logical operators. Shape descriptors including height, width and area thresholds removed any small or oblong non-fish shaped objects from the candidate set. A hue-based threshold was used to filter out any false positives generated by background such as marine snow or clouds of sediment, which had different colour characteristics than target sablefish. Thresholds for merging and noise detection were empirically determined by evaluating results for the experimental database. The output of the Detection step was a binary image representing the segmented fish candidate regions.

The third module (Tracking and Counting) used motion analysis to track the fish candidates and count them. A fish was assumed to enter and leave the frame at a boundary and to move on a connected path, sometimes stopping on the way. The tracking system matched fish through their motion between successive frames. This counting method could detect both unoccluded and partially occluded fish present in the frame. Note that the refinement of the algorithm did not incorporate a machine learning element, but was done by human evaluation of the results and subsequent improvement of the techniques used. To evaluate the algorithm's performance, it was tested on 100 randomly selected videos from the dataset for which the fish were counted manually and compared with the output of the algorithm.

DATA ANALYSIS AND COMPARISON

Data from all groups were matched using the date and time information contained in the metadata. Results from Students and Crowd were automatically recorded in the ONC common database with the accompanying metadata following international ISO 19115 standards. Each annotation is associated with a UserID, the video acquisition and annotation dates and times, and a set of additional metadata (e.g. metadata associated with the instrument, the observatory, the type of data). In the case of the Expert and the Algorithm, data were locally saved on a hard drive and each count was associated with the original video filename that includes the observatory location, type of camera, and date and time of acquisition, allowing for subsequent data combination.

For the Crowd annotators, three groups were identified: the 'Total' Crowd included all data from all participants (503 individuals), the 'Novice Crowd', included data from the first 100 annotated videos of all users, and the 'Advanced Crowd' included videos 101 and higher for all users. An analysis comparing the percentage of correct answers with the number of video processed showed that above 100 videos watched ('Advanced Crowd'), with few exceptions, the percentage of correct counts remained above 70% (Fig. 2a). Only 6·5% of all observers (i.e. 33 individuals) annotated more than 100 videos. Fish classification results for the three different groups of human operators plus the Algorithm were compared considering only videos screened at least once by

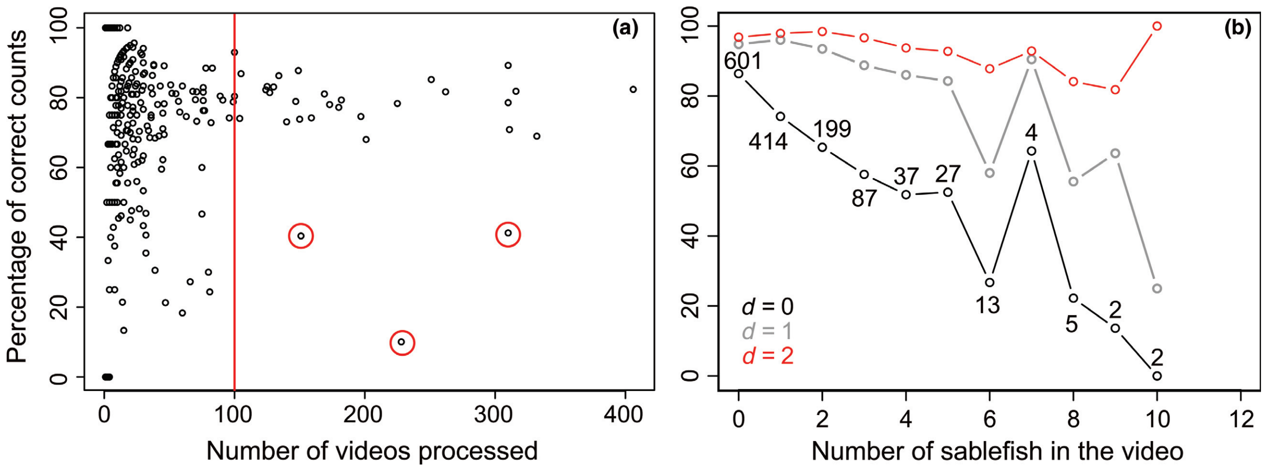

Fig. 2. (a) Percentage of correct counts in relation to the number of videos processed for each member of the Crowd. One citizen scientist who annotated more than 1400 videos was removed from the analyses. The red line depicts the 100 annotated videos threshold. Circles in red depict the only three users who annotated more than one hundred videos but obtained <70% correct counts. (b) Percentage of correct counts in relation to the number of sablefish in the video as determined by the Expert (see text for details). '*d*' provides the margin of error tolerated for the absolute difference in number of fish between the expert and each member of the Crowd, and the numbers on the curves indicate the number of videos containing a given number of sablefish. Both graphs were calculated using 1391 videos processed by both the Expert and the Crowd.

Table 1. Group size (*N*), number of times a video was viewed (*Nt*), Wilcoxon paired rank test and Pearson linear correlation coefficient with Expert for each treatment group (i.e. Experts, Students, Novice crowd, Advanced crowd, Total Crowd and Algorithm)

	Students	Novice crowd	Advanced crowd	Crowd	Algorithm
N	60	503	33	503	1
Nt	1–4	1–20	1–8	5–23	1
Wilcoxon signed-rank test					
Data	–	–	–	–	*
Mean	**	**	**	**	–
Median	*	ns	*	ns	–
Pearson correlation coefficient					
Data	0·90*	0·78*	0·81*	0·79*	0·82*
Mean	0·93*	0·93*	0·92*	0·95*	–
Median	0·95*	0·96*	0·94*	0·97*	–
Differences in counts for individual video clips relative to Expert					
No diff (%)	74·1	71	76·2	72·5	62·9
Positive diff (%)	2·6	15·7	12·5	14·7	6·9
Negative diff (%)	23·3	13·3	11·3	12·8	30·2

*Significant at $P < 0{\cdot}001$, **$P < 0{\cdot}0001$. Differences (diff) in counts relative to Expert provide the percentage of counts within each group that are below or above the Expert counts.

all groups. When there were multiple records of sablefish counts for individual videos (Students and Crowd, Table 1), three statistics were considered: the mean, median and larger mode. Sablefish counts from Students, Crowd and Algorithm were assessed in relation to the Expert 'groundtruthing' data using a Pearson's product moment linear correlation coefficient, and a paired Wilcoxon signed-rank tests. These two tests were performed on the raw data (before combining data), as well as on the mean, median and larger mode calculated on each video. Accuracy was determined by calculating the percentage of counts that fit the Expert's, and the percentage of counts above (positive difference) and below (negative difference) the Expert's. For this, within each group and for each video, the difference was obtained by subtracting individual sablefish count from that obtained by the Expert.

In order to test for groups' abilities to detect similar temporal trends and patterns in the dataset, Whittaker–Robinson periodograms were calculated on fish counts for the Expert and Algorithm and the median for the Students and Crowd in order to screen for periodicities in fish abundance data. Period significance was tested by a permutation procedure (Legendre & Legendre 2012). All data analyses were conducted in R language (R Core Team 2015).

Results

In total, 1059 video files were screened by all four groups (Expert, Students, Crowd and Algorithm). Details on group size and the number of times a video was viewed are listed in Table 1. Over the crowdsourcing (Digital Fishers) campaign period, 503 Citizen Scientists, participated in the mission and collectively contributed 14 192 annotations to 1430 videos.

Over 9 months, each video was on average screened by 10 different Citizen Scientists from both the Novice and Advanced Crowds (Fig. 3). When only considering the Advanced Crowd, each video was only screened two or three times on average, similar to the Students group. In terms of annotations, 27 individual Citizen Scientists (5% of the total Crowd) contributed to more than 50% of the total number of annotations, and among them six (i.e. 1%) contributed 20% of total annotations. The most involved Citizen Scientist contributed 10% of the total number of annotations and annotated all videos included in the campaign.

In general, all groups performed well in comparison to data from the Expert and all Pearson linear correlations were significant (Table 1). Results obtained with the mode matched those of the median and are not presented. For all groups, considering the median (or larger mode) value per video clip improved the correlation with Expert data (Table 1). The paired Wilcoxon signed-ranked test rejected the null hypothesis of no difference between Expert counts and each individual group counts except when comparing against the mode/median for the Novice Crowd and the total Crowd. When comparing raw count data, the Students performed best (cor = 0·90) and the Novice Crowd worst (cor = 0·78). However when comparing the different measures of central tendency, two groups of Crowd (Novice and Total) out-competed the Students and the Algorithm (Table 1). The Crowd as a whole performed slightly better than members of the Novice and the Advanced Crowd with respect to mean and median values, while the Advanced Crowd performed better when considering the raw data. This implies that the use of a central statistic for any group of people decreased the influence of mistakes and thus, a higher number of participants help improve the quality of the results.

The Algorithm displayed the lowest accuracy of correct counts for individual clips (62·9%) and the Advanced Crowd

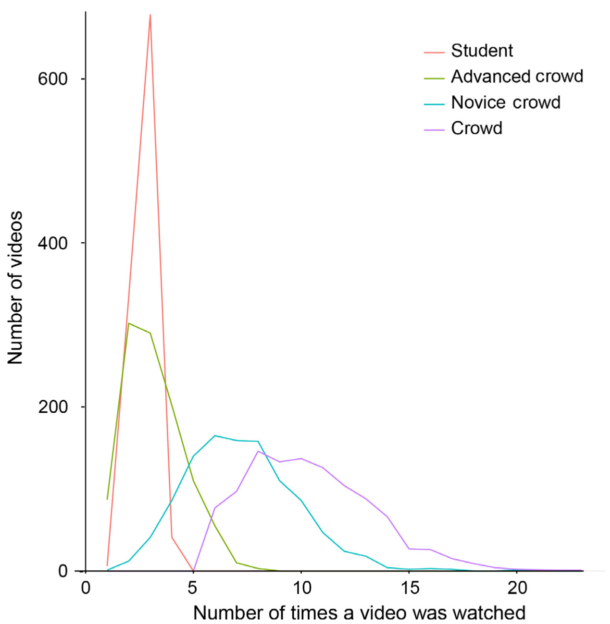

Fig. 3. Frequency distribution of the number of times a video was watched within the different groups.

the highest (76·2%) compared to the Expert (Table 1). The Crowd's accuracy was related to the number of fish in the videos with dramatic increases in 'wrong answers' with increasing numbers of sablefish (Fig. 2b, black line). However, this tendency disappears if we permit a certain margin of error in defining the 'right' answer. Indeed, when allowing for ±2 fish around the real (Expert) value, the percentage of correct answers remains high (Fig. 2b). This latter point is important to consider as missing two fish when only two are present will have greater consequences than missing two when there are 12.

The Algorithm, and to a lesser degree the Students, showed the strongest tendency to undercount fish (30·2 and 23·3% clips undercounted, respectively) relative to the Expert (Table 1). Conversely, the three groups of Crowd tended to overcount (Table 1). Examining count distributions for each video provided insights into the reasons for miscounting. For Students, wrong answers were mostly observed when two fish or more were present in the videos. Missed fish appeared to be those furtively passing in the background or behind other fish, or those for which only a small part enters the field of view, making them difficult to detect. Looking at the Crowd data, several situations were identified: (i) Citizen Scientists tended to overcount as they included fish shadows in their counts; (ii) when a high number of fish passed in front of each other, Citizen Scientists tended to overcount (while Students undercounted); (iii) similarly to Students (but more rarely) undercounting by Citizen Scientists may have been related to missed fish in the shadowed back corners of the field of view, and (iv) in some rare situations where counts were obviously inaccurate, Citizen Scientists may have simply inadvertently hit the wrong key or knowingly entered biased results. It is important to note that this study did not consider miscounting by the Expert.

Despite divergence among the different groups in over- and undercounting, sablefish counts accuracy was >60% for the Algorithm and >70% for the human groups (Table 1). Periodograms calculated for each dataset revealed common periodicities detected by the different groups (Fig. 4). All groups successfully detected a tidal related 12·5 h and 24 h periodicities in the dataset, while a 48 h harmonic was detected by all but the Algorithm. An additional significant periodicity at 64–65 h was identified by the Expert, the Students and the Algorithm.

Discussion

As the deep ocean is increasingly monitored by networks of fixed (i.e. observatories), mobile (i.e. ROVs and AUVs) and semi-mobile (i.e. crawlers) imaging platforms, improving our capacity to extract biological information from underwater imagery is becoming a strategic imperative. Here, we found that human groups (i.e. Citizen Scientists, Students) and an automated computer vision algorithm performed relatively well in counting a single species of fish, using an Expert observer's results (a PhD student) as a benchmark. Until computer vision algorithms become fully competent for such tasks, hybrid solutions that combine machine vision and human

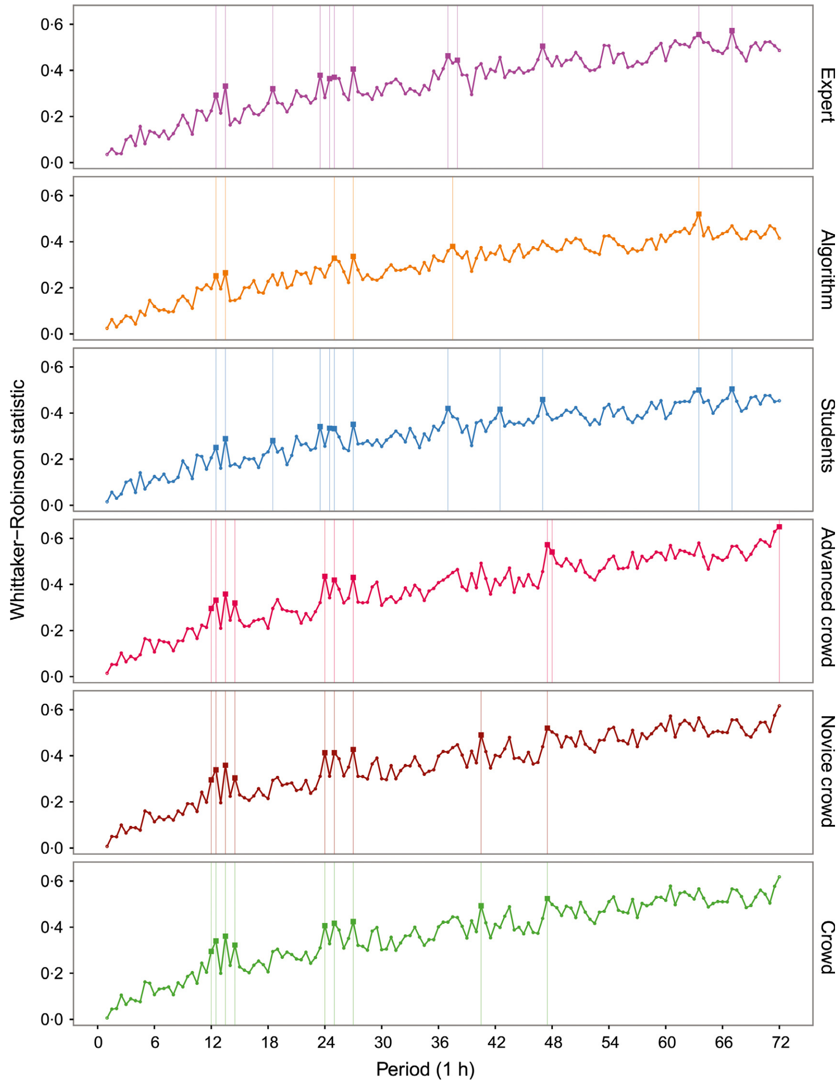

Fig. 4. Whittaker–Robinson periodograms generated from the counts acquired by the different groups. Squares and vertical lines represent significant periodicities. The vertical lines were only drawn to assist in the reading of the period value.

visual discrimination may help reduce the 'image analysis bottleneck' (Gaston & O'Neill 2004; Aguzzi *et al.* 2009). These hybrid solutions will require systematic development and validation, using results from studies such as presented here.

In terms of count accuracy, data from human groups (i.e. Crowd, Students) were nearly equivalent with the highest accuracy (vs. Expert) observed for Students and the Advanced Crowd. Elsewhere, comparisons of marine and terrestrial alpha-diversity data (number of species in a sample/area) obtained by professional scientists vs. volunteers given structured training, have shown that volunteers perform almost as well as professionals (Crall *et al.* 2011; Holt *et al.* 2013). Even for more complicated tasks such as adding measurements to identifications, citizen scientists can provide comparable results to experts (Delaney *et al.* 2008; Butt *et al.* 2013). For other requirements, advanced training may be needed to ensure accurate results. For example, in this study Students outperformed citizen scientists (Crowd) when their results were subjected to periodogram analysis for identification of temporal trends and patterns. They were the only human group that identified all significant periodicities detected by the Expert, corresponding here to the tidal signal (Doya *et al.* 2014). This result is of particular interest for environmental monitoring where detecting trends and events in time series is more relevant than absolute counts. Other studies of citizen science have also observed better performance from highly trained or educated volunteers, highlighting the influence of education on the quality of results (Delaney *et al.* 2008). Note that for this study, advanced citizen scientists were distinguished from novices based on their viewing and annotation experience (more than 100 video clips), a threshold above which citizens had more than 70% correct counts. A high involvement in the project benefitted the user's performance, and could be argued to represent a form of training. On the other hand, the quality of the results can also be a function of the number of volunteers involved. Our study compared 503 citizen volunteers and 60 students against an expert. We obtained the highest correlation with the Expert for the combined results (i.e. median) of the two largest human groups (Novice Crowd and Crowd). Crowdsourcing or 'virtual citizen science' benefits from multiple replications of the same tasks by hundreds or thousands of people, allowing the use of statistics to improve the quality of the results (Wiggins & Crowston 2011; Bird *et al.* 2014; Kosmala *et al.* 2016). Here, the use of the median or mode further increased the strength of the correlation and appeared to be a simple and efficient way to combine large citizen datasets.

In most citizen science studies, volunteers are formally trained in dedicated sessions with professionals, so that their level of expertise is closer to our undergraduate Student category (Azzurro *et al.* 2013). Taking advantage of university classes might provide higher quality results but requires more planning and researcher involvement to establish collaborations, fit projects to teaching programs and priorities, and provide training prior to data processing. In this case, the educational value constituted a priority over data processing. Asking students to complete the task as a course requirement (as we did in this study), could also ensure higher quality results, though outliers, such as the student who systematically annotated zero fish, can also occur. These investments should be weighed against task complexity and potential returns in terms of data quality (Delaney *et al.* 2008). Here, the task to be accomplished was relatively easy and all approaches yielded a valuable solution.

While our results demonstrated that computer vision can yield valuable results for fish population monitoring, the algorithm was the poorest performer when compared against the Expert and the different human groups. The lower performance observed for the Algorithm (compared to Expert, Students and Crowd) can be related to the limitations already identified in Fier, Branzan Albu & Hoeberechts (2014) where fish were camouflaged in the poorly illuminated background, overlapping and occluding each other. It is possible that with additional effort and innovation in the development, the results of the algorithmic method could be improved. Furthermore, the Algorithm results for this dataset might not easily generalize to other seafloor video datasets. Computer vision algorithms are often specific and must be designed to detect and classify particular targets against different background types (Purser *et al.* 2009; Aguzzi *et al.* 2011). Different techniques may be required, for example, to detect and classify marine species of interest in more complex environments where organism densities are high and the background is made of complex 3D biological and mineral structures (e.g. hydrothermal vents or coral reefs). Object detection algorithms perform best in situations of uniform background, such as detecting plankton in the water column (Tsechpenakis, Guigand & Cowen 2007) or benthic animals on soft sediments (Aguzzi *et al.* 2009; Schoening *et al.* 2012). Until computer vision algorithms can overcome these limitations, citizen science and the use of volunteer networks will likely be an important near-term solution for analysing large image datasets from complex marine environments, provided that observer accuracy can be understood, and perhaps improved with training (Dickinson, Zuckerberg & Bonter 2010; Holt *et al.* 2013).

Intermediate, hybrid solutions may also be possible. Ours and other study results suggest that volunteer data can be used to improve machine learning results. For example, in astronomy, where numbers of galaxy images exceed even the processing power of crowds of online citizen scientists, astronomers have successfully used samples of crowdsourced data that had a high degree of internal agreement to train computer algorithms (Kuminski *et al.* 2014). Statistical methods being developed to facilitate the use and validity of citizen science data (Bird *et al.* 2014; Isaac *et al.* 2014) could be used to select subsamples of quality citizen data for machine learning systems. For this, it is essential that any crowdsourcing project includes systematic archiving of metadata in the project development. Here, the quality of the metadata permitted an accurate matching and comparison of annotations from different sources. Our successful combining of results of student and citizen annotations suggests that additional metadata could be generated by an algorithm that would flag videos or images that have been processed by scientists, trained volunteers or citizens, and automatically calculate the median for subsequent statistical comparisons, or to identify high quality datasets for training computer vision algorithms. Another human-machine hybrid approach could involve having volunteers and/or students focus on validating events and trends identified by automated screening systems. This method could enhance participant motivation and improve performance by focussing their attention on higher quality tasks such as verifying abundances or behaviour in specific time blocks identified by the computer processing, rather than sorting long, continuous imagery time series.

Our knowledge of deep-sea ecosystems is limited and fragmented (Ramirez-Llodra *et al.* 2010), at a time when industrial incursions into the deep ocean are increasing with unknown consequences for benthic ecosystems and the planetary support services they provide (Boschen *et al.* 2013; Wedding *et al.* 2013). Remote monitoring that continuously collects imagery is one tool that can be used to document and assess long-term ecosystem change in the deep sea. Realizing the full potential of this technology will require effective solutions for processing massive image datasets to extract relevant biological and habitat information. This study has demonstrated that citizen science, using both crowdsourcing and trained volunteers, together with constantly improving computer vision and machine learning technologies, can contribute to meeting the image processing challenge. In the case of ocean observatories, crowdsourcing, perhaps partnered with algorithms, can help researchers extract trends and events from imagery time series that will improve our understanding of natural variability and therefore our ability to identify anthropogenic impacts. Interactions between science and society have become an important focus for 'big science' programs and infrastructure installations. Citizen science can contribute to developing scientific literacy and informed societal decision-making (Bonney *et al.* 2009). Engaging the public in data analysis will ultimately benefit marine conservation and protection of marine ecosystem services by increasing awareness of our oceans.

Authors' contributions

M.M., M.H., J.A., A.B.A., T.R., S.L., R.M.M., S.K.J.: initial idea and conception of the project. C.D., R.F.: data acquisition. T.R., R.M.M., S.L.: supervision of data acquisition by the students. M.H., A.B.A., R.F.: development of the computer vision algorithm. M.M., J.N.: data analyses. M.M., M.H., J.A., A.B.A., U.F.A., S.K.J.: data interpretation and writing of the paper.

Acknowledgements

The authors would like to thank all the students of the 2012 Biology 335 Ichthyology class at the University of Victoria and the 503 citizen scientists who contributed to this project. We are also grateful to the captain and crew of the R/V *Thomas G. Thompson*, the ROV ROPOS and the Ocean Networks Canada team. J.A. is Theme Leader for the ONC science theme (http://www.oceannetworks.ca/science/science-plan/science-themes/life). We also thank an anonymous reviewer and M. Kosmala for their valuable comments that helped improved the manuscript. Data used in this work were provided by Ocean Networks Canada, a Major Science Initiative recognized by the Canada Foundation for Innovation and supported by the governments of Canada and British Columbia.

Data accessibility

Datasets uses in this study are available for download on the Dryad platform: https://doi.org/10.5061/dryad.98g01 (Matabos *et al*. 2017).

References

Aguzzi, J., Costa, C., Fujiwara, Y., Iwase, R., Ramirez-Llorda, E. & Menesatti, P. (2009) A novel morphometry-based protocol of automated video-image analysis for species recognition and activity rhythms monitoring in deep-sea fauna. *Sensors (Basel, Switzerland)*, **9**, 8438–8455.

Aguzzi, J., Costa, C., Furushima, Y., Chiesa, J., Company, J., Menesatti, P., Iwase, R. & Fujiwara, Y. (2010) Behavioral rhythms of hydrocarbon seep fauna in relation to internal tides. *Marine Ecology Progress Series*, **418**, 47–56.

Aguzzi, J., Costa, C., Robert, K., Matabos, M., Antonucci, F., Juniper, S.K. & Menesatti, P. (2011) Automated image analysis for the detection of benthic crustaceans and bacterial mat coverage using the VENUS undersea cabled network. *Sensors*, **11**, 10534–10556.

Aron, M., Sarrazin, J., Sarradin, P. & Mercier, G. (2010) Elaboration of a video processing platform to analyze the temporal dynamics of hydrothermal ecosystems. AGU Fall Meeting Abstracts, p. 4, San Francisco, CA, USA.

Azzurro, E., Aguzzi, J., Maynou, F., Chiesa, J.J. & Savini, D. (2013) Diel rhythms in shallow Mediterranean rocky-reef fishes: a chronobiological approach with the help of trained volunteers. *Journal of the Marine Biological Association of the United Kingdom*, **93**, 461–470.

Belkin, I.M. & O'Reilly, J.E. (2009) An algorithm for oceanic front detection in chlorophyll and SST satellite imagery. *Journal of Marine Systems*, **78**, 319–326.

Bird, T.J., Bates, A.E., Lefcheck, J.S. *et al*. (2014) Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation*, **173**, 144–154.

Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K.V. & Shirk, J. (2009) Citizen science: a developing tool for expanding science knowledge and scientific literacy. *BioScience*, **59**, 977–984.

Boschen, R.E., Rowden, A.A., Clark, M.R. & Gardner, J.P.A. (2013) Mining of deep-sea seafloor massive sulfides: a review of the deposits, their benthic communities, impacts from mining, regulatory frameworks and management strategies. *Ocean & Coastal Management*, **84**, 54–67.

Butt, N., Slade, E., Thompson, J., Malhi, Y. & Riutta, T. (2013) Quantifying the sampling error in tree census measurements by volunteers and its effect on carbon stock estimates. *Ecological Applications*, **23**, 936–943.

Chuang, M.C., Hwang, J.N. & Williams, K. (2014) Supervised and unsupervised feature extraction methods for underwater fish species recognition. *2014 ICPR Workshop on Computer Vision for Analysis of Underwater Imagery (CVAUI)*, pp. 33–40. Stockholm, Sweden.

Copley, J.T.P., Jorgensen, P.B.K. & Sohn, R.A. (2007) Assessment of decadal-scale ecological change at a deep Mid-Atlantic hydrothermal vent and reproductive time-series in the shrimp Rimicaris exoculata. *Journal of the Marine Biological Association of the United Kingdom*, **87**, 859–867.

Copley, J.T.P., Tyler, P.A., Murton, B.J. & Van Dover, C.L. (1997) Spatial and interannual variation in the faunal distribution at Broken Spur vent field (29°N, Mid-Atlantic Ridge). *Marine Biology*, **129**, 723–733.

Crall, A.W., Newman, G.J., Stohlgren, T.J., Holfelder, K.A., Graham, J. & Waller, D.M. (2011) Assessing citizen science data quality: an invasive species case study. *Conservation Letters*, **4**, 433–442.

Delaney, D.G., Sperling, C.D., Adams, C.S. & Leung, B. (2008) Marine invasive species: validation of citizen science and implications for national monitoring networks. *Biological Invasions*, **10**, 117–128.

Dickinson, J.L., Zuckerberg, B. & Bonter, D.N. (2010) Citizen science as an ecological research tool: challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 149–172.

Doya, C., Aguzzi, J., Pardo, M., Matabos, M., Company, J.B., Costa, C., Mihaly, S. & Canals, M. (2014) Diel behavioral rhythms in sablefish (*Anoplopoma fimbria*) and other benthic species, as recorded by the Deep-sea cabled observatories in Barkley canyon (NEPTUNE-Canada). *Journal of Marine Systems*, **130**, 69–78.

Fier, R., Branzan Albu, A. & Hoeberechts, M. (2014) Automatic fish counting system for noisy deep-sea videos. *Proceedings of Oceans-St. John's, 2014*, p. 6.

Gaston, K.J. & O'Neill, M.A. (2004) Automated species identification: why not? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **359**, 655–667.

Grémillet, D., Puech, W., Garçon, V., Boulinier, T. & Maho, Y.L. (2012) Robots in Ecology: welcome to the machine. *Open Journal of Ecology*, **2**, 49–57.

Hoeberechts, M., Owens, D., Riddell, D.J. & Robertson, A.D. (2015) The power of seeing: experiences using video as a deep-sea engagement and education tool. *OCEANS 2015 – MTS/IEEE Washington* pp. 1–9. IEEE, Washington, DC, USA.

Holt, B.G., Rioja-Nieto, R., Aaron Macneil, M., Lupton, J. & Rahbek, C. (2013) Comparing diversity data collected using a protocol designed for volunteers with results from a professional alternative. *Methods in Ecology and Evolution*, **4**, 383–392.

Isaac, N.J.B., van Strien, A.J., August, T.A., de Zeeuw, M.P. & Roy, D.B. (2014) Statistics for citizen science: extracting signals of change from noisy ecological data (B. Anderson, Ed.). *Methods in Ecology and Evolution*, **5**, 1052–1060.

Kosmala, M., Wiggins, A., Swanson, A. & Simmons, B. (2016) Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, **14**, 551–560.

Kulka, D.W. & Pitcher, D.A. (2001) Spatial and temporal patterns in trawling activity in the Canadian Atlantic and Pacific. *ICES CM 2001/R:02, 57*.

Kuminski, E., George, J., Wallin, J. & Shamir, L. (2014) Combining human and machine learning for morphological analysis of galaxy images. *Publications of the Astronomical Society of the Pacific*, **126**, 959–967.

Legendre, P. & Legendre, L. (2012) *Numerical Ecology*, 3rd edn. Elsevier, Amsterdam, The Netherlands.

Lintott, C.J., Schawinski, K., Slosar, A. *et al*. (2008) Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey ★. *Monthly Notices of the Royal Astronomical Society*, **389**, 1179–1189.

Matabos, M., Hoeberechts, M., Doya, C., *et al*. (2017) Data from: Expert, crowd, students or algorithm: who holds the key to deep-sea imagery 'big data' processing? *Dryad Digital Repository*, http://dx.doi.org/10.5061/dryad.98g01.

Porteiro, F.M., Gomes-Pereira, J.N., Pham, C.K., Tempera, F. & Santos, R.S. (2013) Distribution and habitat association of benthic fish on the Condor seamount (NE Atlantic, Azores) from *in situ* observations. *Deep Sea Research Part II: Topical Studies in Oceanography*, **98**, 114–128.

Porter, J.H., Nagy, E., Kratz, T.K., Hanson, P., Collins, S.L. & Arzberger, P. (2009) New eyes on the world: advanced sensors for ecology. *BioScience*, **59**, 385–397.

Purser, A., Bergmann, M., Lundälv, T., Ontrup, J. & Nattkemper, T. (2009) Use of machine-learning algorithms for the automated detection of cold-water coral habitats: a pilot study. *Marine Ecology Progress Series*, **397**, 241–251.

R Core Team (2015) *R: A Language and Environment for Statistical Computing*. R Development Core Team, Vienna, Austria. Available at: http://www.R-project.org/ (accessed 18 January 2016).

Ramirez-Llodra, E., Brandt, a., Danovaro, R. *et al*. (2010) Deep, diverse and definitely different: unique attributes of the world's largest ecosystem. *Biogeosciences*, **7**, 2851–2899.

Roy, H.E., Pocock, M.J.O., Preston, C.D., Roy, D.B., Savage, J., Tweddle, J.C. & Robinson, L.D. (2012) Understanding citizen science and environmental monitoring. Final report on behalf of UK-EOF.

Schettini, R. & Corchs, S. (2010) Underwater image processing: state of the art of restoration and image enhancement methods. *EURASIP Journal on Advances in Signal Processing*, **2010**, 1–15.

Schoening, T., Bergmann, M., Ontrup, J., Taylor, J., Dannheim, J., Gutt, J., Purser, A. & Nattkemper, T.W. (2012) Semi-automated image analysis for the assessment of megafaunal densities at the Arctic deep-sea observatory HAUSGARTEN. *PLoS ONE*, **7**, e38179.

Silvertown, J. (2009) A new dawn for citizen science. *Trends in Ecology & Evolution*, **24**, 467–471.

Tsechpenakis, G., Guigand, C. & Cowen, R.K. (2007) Image analysis techniques to accompany a new *in situ* ichthyoplankton Imaging system. *OCEANS 2007 – EUROPE*, vol. 1–3, pp. 438–443. IEEE, New York, NY, USA.

Tunnicliffe, V. (1990) Observations on the effects of sampling on hydrothermal vent habitat and fauna of Axial Seamount, Juan de Fuca Ridge. *Journal of Geophysical Research: Solid Earth (1978–2012)*, **95**, 12961–12966.

Vishwakarma, S. & Agrawal, A. (2012) A survey on activity recognition and behavior understanding in video surveillance. *The Visual Computer*, **29**, 983–1009.

Wedding, L.M., Friedlander, A.M., Kittinger, J.N. *et al.* (2013) From principles to practice: a spatial approach to systematic conservation planning in the deep sea. *Proceedings of the Royal Society B*, **280**, 20131684.

Wiggins, A. & Crowston, K. (2011). From conservation to crowdsourcing: a typology of citizen science. *Proceedings of the 44th Annual Hawaii International Conference on System Sciences,* Koloa, HI, USA.

Woodward, G., Dumbrell, A.J., Baird, D.J. & Hajibabaei, M. (2014) Big data in ecology PREFACE. *Adavances in Ecological Research, Vol 51: Big Data in Ecology* (eds M. Woodward, G. Dumbrell, A.J. Baird & D.J. Hajibabaei), pp. IX–XIII. Elsevier Academic Press Inc., San Diego, CA, USA.

Zafeiriou, S., Zhang, C. & Zhang, Z. (2015) A survey on face detection in the wild: past, present and future. *Computer Vision and Image Understanding*, **138**, 1–24.

Received 14 November 2016; accepted 17 January 2017
Handling Editor: Nick Isaac

Supporting Information

Details of electronic Supporting Information are provided below.

Video S1. Example of video from the project dataset recorded in Barkley Canyon (British-Columbia, Canada, using the Ocean Networks Canada Observatory.

Fig. S1. Ocean Networks Canada's annotation system used by the students to count the number of Sablefish in the videos (http://dmas.uvic.ca/SeaTube).

Fig. S2. Tutorial provided to the Crowd participants through the web interface Digital Fishers (http://dmas.uvic.ca/DigitalFishers).

Fig. S3. Summary of automated analysis method to detect fish in the Barkley Canyon videos recorded by the Ocean Networks Canada observatory.