Preliminary Experiments with the Fish4Knowledge Dataset

R B Fisher
University of Edinburgh
Email: rbf@inf.ed.ac.uk

B J Boom
University of Edinburgh
Email: bboom@inf.ed.ac.uk

P X Huang
University of Edinburgh
Email: s1064211@sms.ed.ac.uk

***Abstract*—This paper presents some of the experiments possible when using a very large biological observation dataset, such as the Fish4Knowledge dataset, which was acquired over 1000 days, at 12 hours a day and using 9 undersea cameras. 24 different species were recognized. Each day's observations vary considerably, but use of the large dataset allows trends to be observed. Key results are 1) that there is only little variation in fish observation through the daylight hours, and 2) that typhoons only temporarily disrupt the abundance measures.**

I. Introduction

A previous paper [2] introduced much of the technology used in the Fish4Knowledge project. The project has now completed data collection and the initial data analysis. This paper presents some of the questions that can be answered with the data that was acquired and demonstrates how they might be answered. To answer these questions, the main Fish4Knowledge database is used, which contains processed results from 455993 videos, each of 10 minutes duration.

What is shown is that, in spite of the noisy nature of the data, there are some trends, and also unexpected non-trends. For example, the data will show that typhoons are only temporarily disruptive in terms of the abundance (or at least number of observations). Secondly, there does not seem to be much variation in number of fish observed per hour over the daylight hours.

The following sections give the details of the acquired data, the methodology of the analysis and discuss issues that arose during the analysis.

II. Data

The data analysis presented in this paper is based on the full Fish4Knowledge dataset. The video system recorded 524086 10 minute video clips using 9 cameras at 3 locations. Four cameras were in the bay outside the Taiwan Hengchun Nuclear Power Plant on the south coast, three were in HoBiHu harbor, in Kenting, and two were on the coast of LanYu Island, which was about 50 km southeast of Taiwan.

The raw resolution of the videos was 320x240, at about 5-8 frames per second. However, over half of the videos were recorded at 640x480, and some at 24 fps. This upsizing was created by interpolation and not from higher quality video data. The lower spatial and temporal resolution was beneficial for the storage of the approximately 90 thousand hours of video in about 200 terabytes of compressed storage. The major negative consequence of the capture schema was the slower frame rate, which meant that a fish could travel a considerable distance in the image between frames. This could therefore introduce tracking errors when multiple fish were in the field of view. The lower spatial resolution probably reduced the performance of the recognition algorithms somewhat as well.

TABLE I. Summary of Videos Processed by the Recognition Algorithm (see text for details)

Type	Processed	Total	Percent(%)
Algae	49165	49370	99.58
Blurred	181757	181965	99.89
ComplexScenes	37401	37404	99.99
EncodingProblem	39920	108140	36.92
HighlyBlurred	65025	65024	100.00
Normal	76465	75806	100.87
Unknown	6176	6171	100.08
TOTAL	455993	524086	87.01

Of the 524086 videos, all were processed by the fish detection and tracking algorithms [6], [7]. The first part of the detection process is an analysis of the quality of the videos, using a learning-based classifier. Columns 1 and 3 from Table I show the number of videos estimated to be in each quality category, and also the number from each category that were processed by the recognition algorithm. For a variety of operational reasons, a few videos were processed more than once, leading to some processed percentages being slightly more than 100% and others being less than 100%. Some of the operational reasons were the upgrading of algorithms as the project proceeded and storage of videos at different resolutions. In theory, one could have rerun the analysis using the final algorithms once the project development was finished, but in fact the data analysis used on the order of 400 core years of processing. This implied there is a small amount of error and inconsistency in the recorded database. A machine-learning algorithm was used to assess the quality of the video, and is also used to adjust the parameters of the fish detection algorithms and to prioritize the processing of the videos [7].

Of particular interest was the unexpectedly low percentage of videos that were in the Normal category (14%) and the unexpectedly high percentage of videos that were in the EncodingProblem category (21%), which produced random artifacts in the video, many of which could become false fish detections. The other categories that were classified were Algae (algae on the camera lens, 9%), slightly Blurred water (35%), HighlyBlurred water *e.g.,* after a storm (12%), ComplexScenes where there was much plant and illumination activity in the background (7%), Unknown (1%) and NotSet (0.01%).

The feature extraction stage of the recognition algorithm was considerably slower (implemented in Matlab) and so not all videos were processed by the recognition algorithms [3], [4]. In the end, only 455993 (87%) of the videos were processed. Only about 37% of the videos that were assessed to have EncodingProblems were processed, because they were introducing too many false positve fish detections.

By the end of the project, the detection process [6] detected approximately 1.44 billion individual fish instances of a sufficiently large size (50*50 pixels) that recognition is possible. The detected fish were then linked across video frames [8] to produce approximately 145 million tracked fish. The pixels inside the extracted contours were the inputs into the fish recognition process [3], [4].

The detection process is greatly affected by the illumination conditions on the background, and also any movement of background material, typically seaweed. A consequence of this is a high false detection rate over the full video set. Some experiments ([4] Ch. 5) on randomly selected video clips show that false detections can be on the order of 68%. The recognition algorithm has an ‘unknown or bad detection’ rejection mechanism that eliminates many of these false detections, which rejected 74% of the false detections in the experiment. This reduces the number of tracked and recognized valid fish to approximately 81 million. Furthermore, when the duplication of videos at different spatial resolutions is taken into account, this reduced the number of videos and trajectories to 282048 videos and 57.4 million trajectories. Videos classified as having encoding errors were eliminated, as were videos from 18:00-19:00 (as the recording for these times was highly incomplete). Fish classified by the recognition algorithm as being not one of the 24 trained species or non-fish were rejected, leaving 261751 10 minute video clips (43625 hours of video) and 27.4 million trajectories. All results presented below were based on this final set.

The recognition algorithm was trained using manually produced ground truth [1] of the top 24 species. This is a significantly unbalanced data analysis problem, where the most common species (*Dascyllus reticulatus*) was on the order of 1000 times more numerous than the 24$^{\text{th}}$ species. Altogether, these top 24 species represented 99.7% of fish observed in the groundtruth dataset.

Of the 27.4 million analyzed trajectories, the most commonly recognised species were: *Dascyllus reticulatus* (47% of the dataset), *Scolopsis bilineata* (7%), *Plectroglyphidodon dickii* (11%), and *Amphiprion clarkii* (9%). In the full dataset the ratio of *D. reticulatus* (most common) to *Pempheris vanicolensis* (least common) was 5585:1.

The recognition algorithm performance on the ground truth, when considered over all fish in the top 24 species, averaged 97% correct [4]. However, given the imbalance in the species, we also calculated the average correct recognition rate of each of the 24 species and then averaged these together. In this case, the average recognition rate was 75%, which gives a measure of performance for all species, not simply the most numerous. However, since the real dataset is unbalanced, the 97% correct recognition rate suggests that the performance on the whole video database is also good.

The final comment to make here about the data is the size of the database. The main data storage and processing was done at the Taiwan National Center for High-performance Computing. The 524086 video clips required about 206 Tb of disk storage, and the full SQL database required about 400 Gb to store the details of the detected and recognized fish. See [2] for more details about the hardware, system, video and result database.

III. Analysis

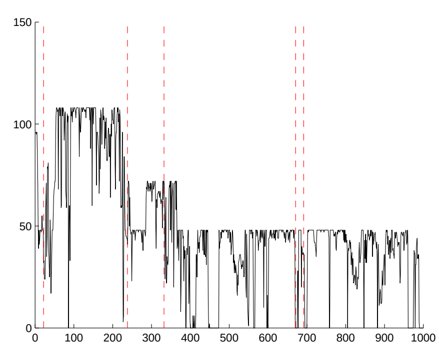
Rather than work with the 400 Gb of individual fish detections, trackings and species classifications, summary tables were used in the analyzes presented below. The two summary tables are:

1) **VideoSummary** - the number of 10 minute video clips recorded as a function of 1000 days (from (1 Oct 2010 - 26 June 2013) $\times$ 12 one hour time slots (from 6:00 to 17:59) $\times$ the number of cameras (9) $\times$ the different video quality categories (6). The video data array is `VideoSummary(1000,12,9,6)`.
2) **FishSummary** - the number of recognized fish (trajectories), indexed as for `VideoSummary`, with an additional index of 24 species. The fish data array is `FishSummary(1000,12,9,6,24)`.

One interesting factor was the nearby passing of 5 typhoons (sustained wind speed of at least 118 km/h) during the observation period: Megi (Oct 22 2010, west of Taiwan, date=22/1000), Songda (May 26 2011, east of Taiwan, date=238/1000), Nanmadol (Aug 28 2011, over Taiwan, date=332/1000), Saola (Aug 1 2012, over Taiwan, date=671/1000), and Tembin (Aug 22 2012, over Taiwan, date=692/1000). These dates are plotted in the figures below with a vertical red dashed line on the date-based plots. The results below show that the typhoons had only short term effects on the number of fish observed.

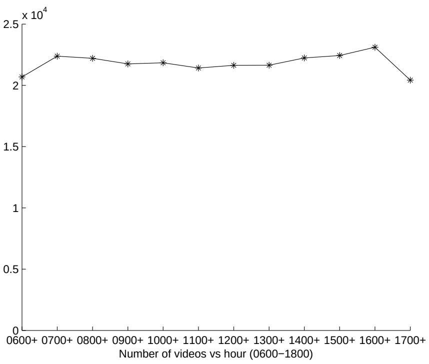
These two summary arrays now allow investigations of a number of questions. Some background questions about cameras and videos are considered first:

1) **What are the number of camera-hours recorded per day?**

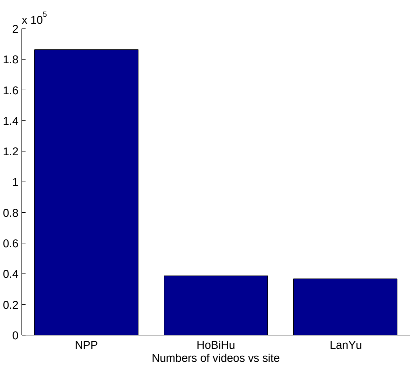


A camera is active if it recorded a video on that day. A camera-hour is counted if any one of the six 10 minute clips was recorded. The data is aggregated over all times and quality levels. The plot above shows: 1) the HoBiHu cameras were active only through day 231, and the LanYu cameras only through day 365, but the NPP cameras were active almost every day. 2) Typhoons caused only short term loss of data. 3) If the cameras were active, we achieved almost a full day of acquisition.

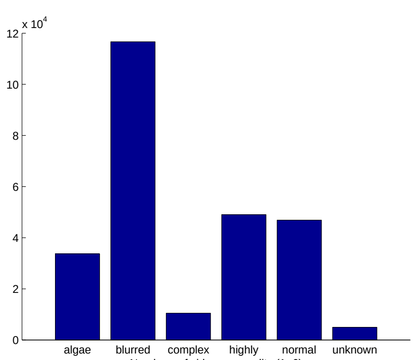
2) **What are the number of videos recorded per hour?** These are aggregated over all sites and quality levels. The plot below shows approximately the same number of videos were recorded at each hour, with a slight reduction at dawn and dusk.



3) **What are the number of videos per site?** The plot below is aggregated over all days, times and quality levels, and also aggregates (1) the four cameras at the Nuclear Power Plant, (2) the three at HoBiHu harbor and (3) the two at LanYu island. The plot is as expected, given the successive loss of cameras over the 1000 days.

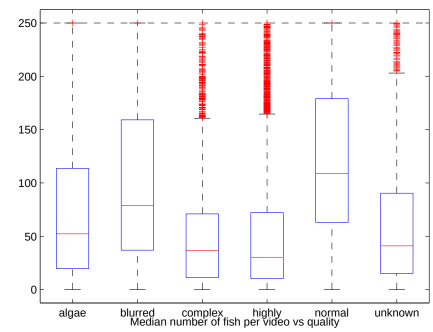


4) **How many videos were analyzed for each quality classification?**



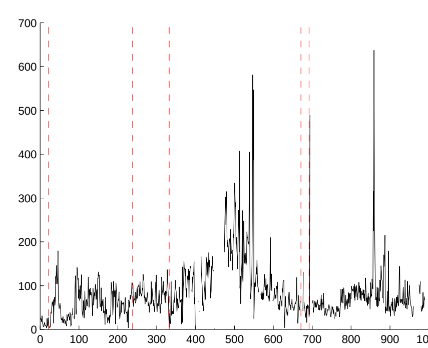
The plot above was aggregated over all days, times, sites and species. The quality measures (horizontal axis) were 1: algae on the lens, 2: slightly blurred water, 3: complex changing backgrounds, 4: highly blurred water, 5: normal, and 6: unknown. The plot shows that many of the videos were recorded with slightly blurred water.

5) **How many fish were detected per video for each quality classification?** This plot shows the median number of fish detected in a video for each of the six quality settings. It shows that most fish are detected with normal and slightly blurred videos, as one can expect. This assumes that the same distribution of environmental effects are experienced for each camera condition. This is probably not the case with the highly blurred water, which is likely to occur during storms, and thereby also affect the fish.



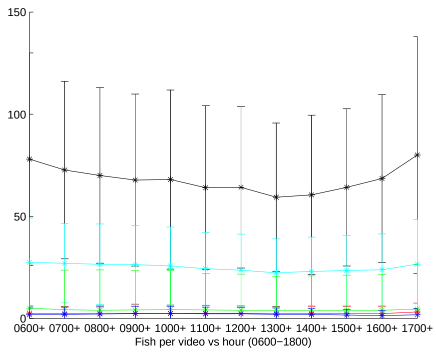
Some questions about the fish abundance over the whole observation period are now considered:

6) **What are the numbers of fish observed per video per day?** The plot below is aggregated over all times, sites, qualities, and species and shows the median number of fish divided by the number of videos captured on that day. The plot shows that there is a lot of variability in the observations, and there are some noisy measurements (probably due to undetected compression artifact failures) at days 547 and 859. The bulges in the data near days 503 and 866 (February 15 2012 and 2013) led us to wonder if there was seasonal effect. The data near day 138 (15 February 2011) are slightly suggestive of a repeating seasonal peak.

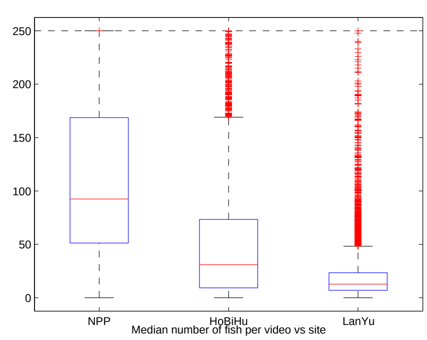


7) **Does the number of fish per video vary according to the time of day?** The plot below is aggregated

over all days, cameras, qualities and species and is the median number of fish per hour, with a robust estimation of the ±1 standard deviation error bars. The black curve is over all species, cyan: *D. reticulatus*, red: *S. bilineata*, green: *P. dickii*, and blue: *A. clarkii*. The plot shows a slight increase in the median value of the total count at dawn and dusk, but the variances are quite large. It's hard to make any conclusions for the individual species.



8) **Does the number of fish per video vary according to site?** The plot below was aggregated over day, time, quality and species, and show the median number of recognized fish per video, with a robustly estimated 1 standard deviation error bar. Cameras 1, 2, 3, and 6 were at site NPP (site 1); cameras 4, 5, and 7 were at site HoBiHu (site 2) and cameras 8 and 9 were at site LanYu (site 3). The plot suggests greater fish abundance and variability at the NPP site.



IV. Discussion

The Fish4Knowledge project has collected one of the largest video and specialized image databases in the world, based on advances in target detection, tracking and class-based recognition. The quantity of the data has helped improve the understanding of tropical coral reef fish near Taiwan.

The collection of the data over such a long time period has also exposed many issues arising in such a challenging project, the most important of which is of the quality of the data. The outdoor ocean environment near the surface is visually difficult, with the most problematic effect being the dramatically changing illumination on the background because of the caustics arising from the changing ocean surface. Varying quality of the water media and plant life in the background also causes problems. A further issue is the occasionally defective video arising from the many technical challenges in acquiring off-shore video, cabled communication to shore, heavy compression and then use of telephone line based upload to the supercomputer center. Cameras and facilities degrade and algae grows on the lens. The result of these difficulties is a somewhat incomplete dataset, and which seems to have about a 68% false detection rate, many of which are removed by the recognition algorithms.

Another factor is the time period over which the data was collected and analyzed: the enormous amount of computation required meant that it was not possible to analyze all videos with the final versions of the algorithms. Thus results from videos processed in the autumn of 2012 were not as accurate as those processed in the summer and autumn of 2013. Recoding the recognition algorithms from Matlab to C++ would have helped greatly, but there was not enough project time.

An interesting 'marine ecology' aspect of the dataset is the large percentage (67%) of the total fish observations represented by the top three species mentioned above, which are known to be non-migratory: *D. reticulatus, P. dickii* , and *A. clarkii*. The implication of this is that multiple observations of the same individual are very likely because the fish swims out of the field of view and then returns. Because of the low camera resolution it is difficult to determine if the individual is the same as previously observed. Interestingly, Liu [5] demonstrated that one could cluster individual *A. clarkii* (clownfish) based on their distinctive stripe patterns. From this small experiment using 785 images, we estimate that we might be over-counting resident *A. clarkii* individuals by a factor of 10.

Acknowledgements: This research was funded by the European Commission (FP7 grant 257024) and the Taiwan National Science Council (grant NSC100-2933-I-492-001) and undertaken in the Fish4Knowledge project (www.fish4knowledge.eu).

References

[1] Boom, B J, Huang, P X, He, J, Fisher, R B (2012) Supporting Ground-Truth annotation of image datasets using clustering. Proc. 21st Int. Conf. on Pattern Recognition (ICPR)

[2] Boom, B J, He, J, Palazzo, S, Huang, P X, Chou, H-M, Lin, F P, Spampinato, C, Fisher, R B (to appear) A research tool for long-term and continuous analysis of fish assemblage in coral-reefs using underwater camera footage. Ecological Informatics

[3] Huang, P X, Boom, B J, Fisher, R B (2012) Underwater Live Fish Recognition using a Balance-Guaranteed Optimized Tree, Proc. Asian Conf. on Computer Vision, Daejeon, Korea

[4] Huang, X (2014) Balance-Guaranteed Optimized Tree with Reject option for live fish recognition. PhD Thesis, University of Edinburgh

[5] Liu, X (2013) Identifying Individual Clown Fish. MSc Dissertation, School of Informatics, University of Edinburgh

[6] Palazzo, S, Kavasidis, I, Spampinato, C (2013) Covariance based modeling of underwater scenes for fish detection. Proc. Int. Conf. on Image Processing, Melbourne

[7] Spampinato, C, Palazzo, S, Kavasidis, I (2014) A texton-based kernel density estimation approach for background modeling under extreme conditions. Computer Vision and Image Understanding, 122:74-83

[8] Spampinato, C, Palazzo, S, Giordano, D, Kavasidis, I, Lin, F-P, Lin, Y-T (2012) Covariance Based Fish Tracking In Real-Life Underwater Environment. Proc. Int. Conf. on Computer Vision Theory and Applications (VISAPP), Rome