Ecological Informatics 23 (2014) 83–97

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

A research tool for long-term and continuous analysis of fish assemblage in coral-reefs using underwater camera footage

Bastiaan J. Boom [a], Jiyin He [c], Simone Palazzo [b], Phoenix X. Huang [a], Cigdem Beyan [a], Hsiu-Mei Chou [d], Fang-Pang Lin [d], Concetto Spampinato [b], Robert B. Fisher [a]

[a] School of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh, EH8 9AB, United Kingdom
[b] University of Catania, Viale Andrea Doria, 6, 95125, Catania, Italy
[c] Center for Mathematics and Computer Science (CWI), Science Park 123, 1098 XG, Amsterdam, The Netherlands
[d] National Applied Research Laboratories, No. 7, R&D 6th Rd., Hsinchu Science Park, Hsinchu City, R.O.C. 30076, Taiwan

ARTICLE INFO

Article history:
Received 31 January 2013
Received in revised form 28 October 2013
Accepted 29 October 2013
Available online 11 December 2013

Keywords:
Video analysis
Long-term monitoring
Fish abundance
High-performance computing
Fish-species recognition

ABSTRACT

We present a research tool that supports marine ecologists' research by allowing analysis of *long-term* and *continuous* fish monitoring video content. The analysis can be used for instance to discover ecological phenomena such as changes in fish abundance and species composition over time and area. Two characteristics set our system apart from traditional ecological data collecting and processing methods. First, the continuous video recording results in enormous data volumes of monitoring data. Currently around a year of video recordings (containing over the 4 million fish observations) have been processed. Second, different from traditional manual recording and analysing the ecological data, the whole recording, analysing and presentation of results is automated in this system. On one hand, it saves the effort of manually examining every video, which is infeasible. On the other hand, no automatic video analysis method is perfect, so the user interface provides marine ecologists with multiple options to verify the data. Marine ecologists can examine the underlying videos, check results of automatic video analysis at different certainty levels computed by our system, and compare results generated by multiple versions of automatic video analysis software to verify the data in our system. This research tool enables marine ecologists for the first time to analyse long-term and continuous underwater video records.

1. Introduction

One of the new challenges in today's data-driven world is how to make sense of enormous amounts of data (Kelling et al., 2009). To gain a better understanding of a complex environment such as a coral reef, collecting data for long-term monitoring of these environments is essential. Long-term monitoring of a coral reef environment can however be labour intensive, requiring divers to identify and count the fish species in a certain area (Pattengill-Semmens and Semmens, 2003). A number of disadvantages of the data collected by divers have been discussed in the literature (Hill and Wilkinson, 2004), including that the presence of divers may affect the fish assemblage, and that divers differ in their experience and ability to identify species.

Fixed underwater cameras can be used continuously to record the coral reef environment during the daytime. Compared to diver-collected data, camera collected data avoids some of the disadvantages of diver collected data. For example, fish activities are not influenced by the sensing equipment, the recorded video footage can be reused by multiple interested parties and video footage can be analysed by different kinds of automatic software as well as different marine ecologists. More importantly, continuous recording may capture trends and developments in the environment that may be missing from divers' observations. On the other hand, data collected by underwater cameras also brings new challenges both in creation (Jan et al., 2007) and analysis (Ebner et al., 2009) of this kind of data.

Analysis of this kind of data requires either a lot of human effort (Ebner et al., 2009) or automatic video analysis technologies. Advances in automatic video analysis technologies (Huang et al., 2012a; Spampinato et al., 2010) yields new solutions to address the above challenges. The goal of this research is to develop a system that allows marine ecologists to access and analyse the video content. In this case, the system is able to automatically find and recognise certain fish species in the video. This information is then organized and presented to users with a web interface for further analysis. This allows the users (marine ecologists) to analyse statistical summaries of the fish species count determined by automatic video analysis technologies. Users can create and verify hypotheses based on this data by checking the videos or performing additional diving expeditions. Currently, our dataset consists of video footage collected by up to 10 cameras that have been recording during 12 daylight hours for the last 3 years.

The contribution of this research can be summarized as follows. First, this is the only system that is able to *analyse of underwater video recordings for the presence of fish*, which makes these results no longer only dependent on the work of divers. An advantage of video recording is that the data becomes reusable, it also allows other marine ecologists to analyse and verify the results afterwards. Second, it is the first system that

1574-9541/$ – see front matter. Crown Copyright © 2013 Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecoinf.2013.10.006

gives marine ecologists *a user interface to analyse and explore the output of automatic video processing software*. Third, the amount of analysed data by the system is unique, where we already have around *4 million observations of fish from around 1 year of continuous videos of multiple cameras*.

This paper describes the first prototype that is able to perform the challenging task discussed above. After the related work (Section 2), an illustrative example of the output of the system is given in Section 3 showing the ability to analyse and present new trends in marine ecology. In particular, we focus on three key aspects of the system: (i) data-intensive processing of underwater video footage (Section 4); (ii) fish detection and species recognition (Section 5) and (iii) visualization of the data (Section 6). Evaluation of video/image processing software is presented in Section 7 and an example is given how to verify observations with the user interface in Section 8.

2. Related work

To our knowledge, this is the first research that aims to analyse multiple years of underwater video data as described in the previous section. A number of research lines are relevant to the type of work described here. These research lines include: studies in analysing large ecological datasets; projects that use underwater cameras to monitor certain aspects of the underwater environment; and computer vision methods developed for recognizing fish species.

Large data collections for scientific purpose have received much attention in recent years. Most of these data collections are developed with human-observers inserting data or observations. One of the most well-known projects is Galaxy Zoo[1], where human volunteers can classify galaxies into different shapes. Research on large data collection, more related to ecology is ebird[2] (Sullivan et al., 2009), where volunteers upload their observations to a website, which allows scientists to look at location-based biological patterns of birds using large numbers of observations. Similar projects exist for flora observations (Auer et al., 2011; McGuire et al., 2008), where both projects couple observations to physical locations. To monitor the coral reef, there is a similar project[3] allowing divers to share voluntarily their observations online. As already discussed, human observations in this case share the same disadvantages as in other diver-collected data. All these current systems rely on human volunteers to insert data, while our system is fully automatic.

Instead of using diver observations, video recording can be analysed which avoids some of the disadvantages of diver observations. In Table 1, a comparison between the pros and cons of using video recording and diver observation is given. An overview of underwater camera systems for monitoring this kind of environments is given in (Shortis et al., 2009). Different camera setups are used for fish observations: cameras with and without bait (Watson et al., 2005), different kind of bait (Dorman et al., 2012), stereo vision (Cappo et al., 2006) and high-resolution rotating cameras (Pelletier et al., 2012). Most previous work uses short term video recordings, where our system used video data that continuous monitors the coral reef (Jan et al., 2007). The analyses of videos are often still performed by human observers except for the size of the fish which is usually a combination of human annotation and stereo vision (where 3D depth of the scene is determined with two cameras). A related research topic to fish identification is plankton identification. Plankton is much smaller than fish, so specialized sensing equipment is necessary. Software has been developed to classify up to 10–20 taxonomic classes with an accuracy of around 70–80% (Benfield et al., 2007). Examples of software for plankton classification are Visual Plankton (Davis et al., 2005), PICES (Luo et al., 2004) and ZOOSCAN[4].

Table 1
Table containing the pros and cons of diver observations versus video recording.

	Diver observations	*Video recording*
Fish activities:	Changed due to present of divers	Go back to normal after installation/maintenance of equipment
Mobility:	Divers swim around	Camera often static
Visibility	Larger field of view	Smaller field of view
Time:	Diver are limited by oxygen (hours)	Camera are limited by maintenance (weeks)
Recognition:	Human recognition (most cases better)	Automatic fish recognition
Consistent:	Human has attention span	Very consistent
Repeatable:	Diver observations cannot be verified	Video recording can be double checked by expert

While the literature and software of automated plankton identification are voluminous, software and literature on fish recognition are less common. One of the possible reasons for this might be that fish in a natural environment are more difficult to classify, because the difference between fish species are more subtle and there are more taxonomic classes in comparison to plankton. Another reason might be that no expert equipment is necessary, so for small numbers of fish humans can easily perform the analysis themselves, however for larger datasets this becomes impossible. Automatic fish species recognition has been developed for different purposes, both for commercial applications like fish farming and fishery and for environmental monitoring. There is research on automatically measuring fish size and estimating their biomass using stereo vision (Ruff et al., 1995; Strachan, 1993a; Strachan et al., 1990). Early work in fish recognition (Strachan, 1993b; Strachan et al., 1990) is focussed on fish on conveyor belts and classifies fish based on shape and colour. Classification of fish in aquariums and tanks (Lee, 2004; Toh et al., 2009) is more challenging than classification of dead fish (Larsen et al., 2009). The first research in unrestricted natural environments (Rova et al., 2007) is able to classify between two different species, where Spampinato et al. (2010) classified 360 images of ten different species (which is one of the largest datasets mentioned in literature). Extensions of the work of (Spampinato et al., 2010) are used in this paper for fish detection and tracking and species classification. Until recently, fish recognition software dealt with very small datasets. In this work, the system made more than 4 million observation of fishes in the video recordings (however many of these are resident species so are frequently re-observed). These observations are stored in a database, where a web interface allows different visualization options to explore this data, giving marine ecologists the ability to look at trends in fish count over time (i.e. hours in a day, fluctuations in a year).

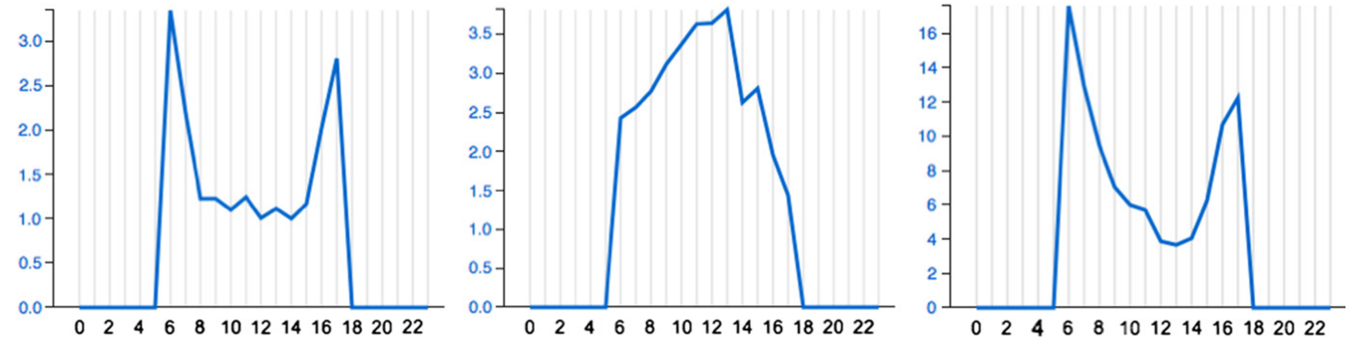
3. Illustrative example of system usage

By using a scenario, we show how this system (webinterface) can be used for instance to explore temporal patterns in fish counts. The system provides users a webinterface http://f4k.project.cwi.nl/data1/ui/ that allows user to select counts of different species over years, hours in week, camera sites, etc. While observed patterns may not have an obvious association with an existing biological/ecological explanation, such information may be useful in providing entry points for marine ecology researchers to conduct further investigation, e.g., in terms of formulating hypothesis and design diving experiments.

3.1. Data exploration scenario

While looking at the counts of different fish species throughout the different daylight hours, we notice that the count distribution of *Chromis*

[1] www.galaxy-zoo.org.
[2] www.ebird.org.
[3] www.reef.org.
[4] www.zooscan.com.



(a) Camera 37

(b) Camera 38

(c) Camera 39

Fig. 1. Average count of Chromis margaritifer per video over time-of-day.

margaritifer shows a unusual pattern different from that of other species: there is a peak at around 7 am and a peak at around 17 pm, while during the day the counts are generally lower. Similar patterns have been found in the counting data generated by different video analysing software (i.e., different versions of fish detection, tracking and recognition algorithms, see Section 5). However, this pattern is not consistent across videos captured by different cameras. Fig. 1 illustrates the counting results for different cameras, and Fig. 2 shows the scenes captured by each camera.

This observation was not confirmed to associate with a known activity pattern of this species when consulting the marine ecologists/biologists from our project advisory board. Nevertheless, such information enables the discovery of possibly interesting phenomena that may worth further investigation. For instance, given the observation from the data, it may be interesting to examine whether the counts of this particular species vary with respect to spatial *and* temporal changes (i.e., time-of-day).

In summary, our system allows users to explore patterns from the processed video data, although a correct interpretation of an observation may require substantial further investigation, e.g., comparing observations from videos to diving experiments, and observations may need to be associated with more information than what is recorded by the cameras, e.g., tidal state and weather conditions.

4. Overview of the research system

4.1. Challenges

Developing a system to analyse large volumes of video recording is a challenging task. Researchers in computer science (High Performance Computing (HPC), computer vision, human computer interaction) have joined efforts in order to create this tool where the main challenges are in:

- *The processing and storage of large volumes of video footage*: In this case, the High Performance Computing facility in Taiwan is used for the storage of videos (200 TB), processing of videos (with up to 1000 CPUs) and the storage of processed data. Without these facilities, it is impossible to analyse the current underwater videos recording.
- *Automatic detection and species recognition in video footage*: Computer vision in uncontrolled conditions is often difficult. This is however complicated by algal growth on the lenses, movement of vegetation, complex coral backgrounds, and time-varying illumination patterns caused by surface waves. Because the automatic fish detection and species recognition will not be perfect in these videos, our system provides marine ecologists a web interface to check the results under varying levels of uncertainty in the results. First, certainty scores between (0…1) allow users (marine ecologists) to select levels of confidence given by the video/image processing software. Second, this system can use various different video/image processing software modules for both fish detection and species recognition. This allows users to confirm observations, by verifying them with for instance a second software module. If multiple software modules observe the same trends, there is a higher probability that the observed trend is not due to systematic errors in the video/image processing software.
- *Dynamic usage of a research tool*: Given the large volumes of collected data, it is not obvious which observations are interesting for users of the system. Although marine ecologists indicated that they are interested in monitoring the abundance[5] (counts) of certain fish species over large periods of time, they are not sure what results to expect. The interface has to give them the possibility to explore the processed data. Ecologists can search for trends in the data, verify these trends by looking at videos themselves, checking if different software modules for fish detection and species recognition show similar trends, etc.

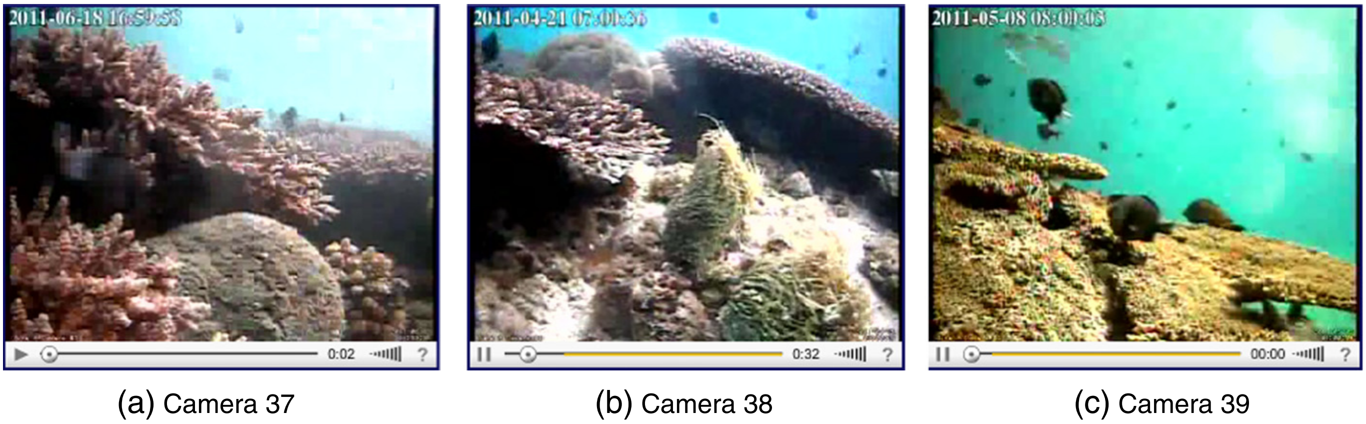
4.2. Software Modules

Given the above challenges, one of the important issues of this research tool is that it allows for flexibility. To achieve this flexibility, the system consists of several software modules and data storage facilities which are connected as shown in Fig. 3. In the case of the Video/Image Processing (VIP) modules (fish detection/tracking and species recognition), different methods can be used for the same task (fish can be found with GMM or ViBE algorithm, see Section 5.2 for more information). This gives the marine ecologists the ability to verify their findings with different methods in case one method gives systematic errors. All the software components communicate by means of the database definitions of the VIP database (Fig. 3). The fish detection/tracking software find the fish position in the frame and follows the fish through consecutive frame, giving a “fish trajectory[6]” in the analysed video. This fish trajectory is stored into the VIP database, based on this information, the fish recognition software determines which species is observed in the fish trajectory. Both fish detection and recognition modules also store a version number, which allows ecologists to determine which software module produced the data making it possible to verify the data with other software modules. A user interface has been developed for analysing the data currently collected. Summary tables are necessary to make the user interface fast enough to deal with the large volumes of data.

In order to run this software in a High Performance Computing (HPC) environment, a simple workflow program (the red arrows in Fig. 3 show the workflow) has been developed that keeps track of all

[5] The fish counts in our system cannot be directly linked to current biological measure for abundance. Cameras have a small field of view, allowing fish to swim in and out the field of view where the system counts these fish multiple times, divers are however unlike to count the reobserved fish. However estimations of the abundance are possible based on comparing observations of diver and camera in the same region.

[6] Obtaining fish trajectory allows us to improve the species recognition because classifying the species is in some frames easier than other frames, because some distinctive features are not always visible. Without trajectory information we are counting the fish in every frame it is visible, creating huge biases towards species who stay in the same place/swim slowly.



(a) Camera 37 (b) Camera 38 (c) Camera 39

Fig. 2. Scenes from different cameras.

the processing by means of the *processed_videos* table. The goal of the workflow is to run the VIP software modules on all recorded video footage, later versions of our system will allow the marine ecologist to request the VIP software to be run on specific video footage. The workflow software is able to execute the VIP software modules, which in turn report the status in the *processing_video* table (for instance if they are running, crashed or finished). The workflow executes the VIP software modules in the HPC environment using the Load Sharing Facility (LSF) to schedule the executable on multiple CPUs (up to 1000 CPU). The dependencies of the software modules are simple, namely that species recognition depends on the fish detection component. The workflow checks automatically for videos that have been processed with the fish detection software module.

4.3. Video and processed data

Currently, our dataset consists of footage from multiple cameras, which have been recording video during 12 daylight hours for the last 3 years, where an earlier versions of the camera site and recording systems is described by (Jan et al., 2007). In Fig. 4, some of the typical video recordings are shown. There are up to 10 cameras which recorded during 12 daylight hours for the last 3 years. There are no overlapping fields of views between the individual cameras, making estimation of the fish size impossible. We use a typical CCTV underwater camera with a focal length of 3.6 mm. The video resolution of older videos is 320 × 240 with 5 frames per second. Due to improvements in the setup higher resolution videos and frame rates are available (640 × 480 up to around 20 frame per second). The videos are saved in 10 minute clips and are stored NAS (network-attached storage) of 200 TB. As of 1 January 2013, we ran the fish detection on 41,815 video clips of 10 min, and the fish recognition on 30,048 videos of 10 min of the 525,450 video clips. In these videos, 6,816,473 individual fish are detected and species recognition is performed on 4,897,786 individual fish[7]. An individual fish appears on average in 11 frames, however many of these are resident species so they are frequently re-observed by the cameras. After running the fish detection software on the videos, the data is added to the tables *fish* and *fish_detection*, where a single fish (*table:fish*) is visible in multiple frames (*table:fish_detection*) allowing us to save the fish trajectory information. The fish recognition software reads this information together with the videos and inserts in the *fish_species* table the species of each detected fish. All these tables also contain a certainty field, which the video/image processing components use to indicate with a value between 0 and 1 how certain they are that their decision is correct. The table, where video/image processing results are stored, also contains a link to the software component table, which allows the users to check which software produced the data.

5. Video and image processing

5.1. Introduction

Video and image processing software allows the system to automatically analyse the underwater videos. Lots of research in computer vision has been performed for video surveillance in the human/urban environment. However, in an unconstrained underwater environment a few difficulties arise that are not present in the human/urban environment. For instance:

- *Video limitations*: Communications between the underwater cameras and the storage servers can be troublesome, where decoding artifacts occur because the data transmission sometimes fails, because of distance between the remote location of the cameras and the storage facilities (due to cable length, power of signal). This makes it often impossible to fully exploit the capabilities of the capture devices (e.g. highest resolution and frame rate).
- *Water cleanness and clarity*: Environmental phenomena (such as storms) and strong currents can reduce the visibility underwater, thus limiting the view distance and the definition even of close objects.
- *Lens cleanness*: If the lenses are left underwater for a long period, lenses inevitably suffer from the growth of algae on them, whose effects can vary from a strong blurring of the image (or part thereof) to the complete obstruction of the field of view. Human intervention is periodically needed to prevent or solve such problems.
- *Lighting*: Time-varying illumination patterns caused by surface of the waves is a difficult problem to tackle, because the frequency of the waves is not constant in coastal areas. Further, the sun and cloud positions vary over time.
- *Fish motion pattern*: When dealing with people or vehicles, it is possible to build an a-priori motion model which allows one to predict the behaviour. However, defining such a model is much more challenging for fish, since their typical erratic motion pattern makes it very difficult to predict their next movements and therefore to track the targets. This is partly because of the unpredictable effects of local ocean currents and partly because fish have many more degrees of freedom in the water compared to humans or cars.

To handle these problems, specific algorithms have been devised to run on underwater videos and achieve similar performances as on more common environments. The following subsections will introduce methods for fish detection, tracking and species recognition that have been used in this system.

[7] As of 30 July 2013, we ran the fish detection on 525,450 video clips of 10 min, and the fish recognition on 118,281 videos of 10 min, resulting in 123,466,562 individual fish detections.

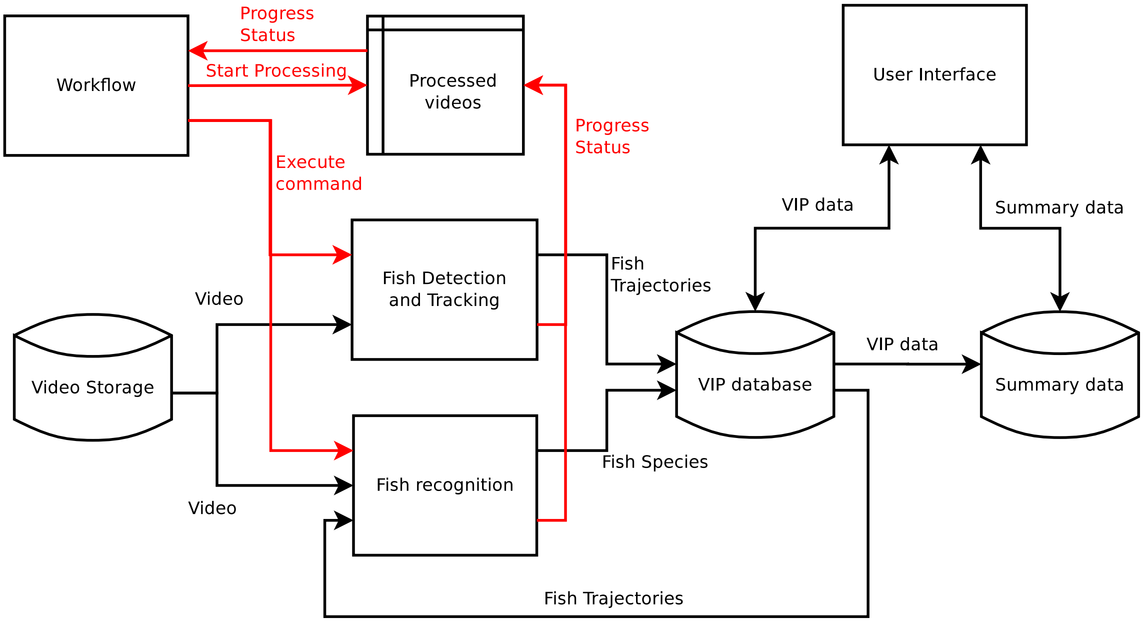


Fig. 3. Schematic overview of the research tool for analysing fish assemblages with video data. The black arrows show the dataflow in the system, while the red arrows show the workflow of the system. This system finds the "fish trajectory" (positions of single fish in the multiple frames) in the video and given the fish trajectories determines the most likely fish species. This information is all stored in a Video/Image Processing (VIP) database. The user interface queries the summary data from this VIP database allowing marine ecologists to analyse this information.

5.2. Fish detection

Most of the existing literature for automatic video and image analysis in underwater environments is focused on fish recognition (Spampinato et al., 2010), whereas few detection approaches exist which are specifically designed for underwater videos. However, many algorithms have been proposed which tackle some of the problems described in the previous section and therefore were found suitable for our needs.

In this project, we selected and improved a few approaches from the state of the art which looked promising for our purpose, either because they had been largely and successfully applied in similar applications or because they targeted aspects of the problem which are particularly relevant to the underwater case, and integrated them into the detection/tracking framework. The well-known Gaussian Mixture Model (**GMM**) (Stauffer and Grimson, 1999) and its Adaptive Poisson Mixture Model (**APMM**) variant (Faro et al., 2011) has been implemented, as an example of the application of mixture-based algorithms. Such methods model for each pixel the distribution of the intensity values which typically describe the background. Mixture-of-model approaches can potentially converge to any distribution, given enough observations; however, the computational cost grows exponentially with the number of models. The GMM approach uses Gaussian distributions to model the background, and is able to deal with multi-modal backgrounds, although not as well with frequent or abrupt lighting changes. APMM employs Poisson distributions, based on the consideration that the intensity of pixels is Poisson-distributed, and should handle illumination variations better. An adaptation to the underwater domain of the approach (**IM**) described in (Porikli, 2005) has been also developed; it specifically deals with sudden illumination changes, by separating, in each frame, the reflectance component (which is a static element of the scene) from its illumination component (which varies depending on the current lighting conditions). The background model is then computed as a temporal median of these two components. The **ViBe** algorithm (Barnich and Van Droogenbroeck, 2011) does not presume the existence of any underlying statistical distribution and employs random element and time sampling when updating the background, resulting in a very efficient yet accurate model. It stores for each pixel a list of its 20 most recent intensity values, and compares the current value in each new frame to the pixel's *history*; a high number of close matches will mark the pixel as background. Finally, as a combination of all previous approaches, an **Adaboost** (Freund and Schapire, 1997) classifier has been trained, which internally employs all of the above-mentioned methods.

All of these algorithms provide as output a motion binary mask where 1's and 0's respectively correspond to foreground and background pixels. Low pass filters are then applied to remove isolated pixels which may be due to noise or misclassifications by the detection algorithm, then contiguous foreground regions ("blobs"), which are likely to represent moving fish and often provide an accurate enough segmentation of the contour from the rest of the scene.

However, due to the instability of the background and the occasional misbehaviour of the motion detection algorithm, a blob post-processing layer is added at the end of the pipeline to filter out bad detections and reduce the number of false positives. The detection post-processing module, described in detail in (Spampinato and Palazzo, 2012a), analyses each blob by integrating a perceptual organization model with intraframe and interframe properties into a Bayesian classification framework, to verify that its shape, texture, motion, structure and segmentation characteristics match those that one would expect of a correctly-identified fish.

5.3. Fish tracking

This section describes the fish tracking algorithm we adopted for the processing of underwater videos. From a marine ecologist's point of view, tracking accuracy is extremely important in the monitoring of a fish assemblage, since it is directly linked to its size, which makes the choice of the tracking algorithm particularly delicate. Moreover, trajectory extraction is also the first step for a high-level activity analysis module, which will be addressed in the future for the study of fish behaviour.

However, the same environmental problems which affect fish detection have to be taken into consideration in the choice and development of a tracking algorithm:

- Together with motion properties, appearance is the most important feature for distinguishing the targets from each other. However, fish represent a very difficult case to handle, because individuals belonging to the same species are practically identical.

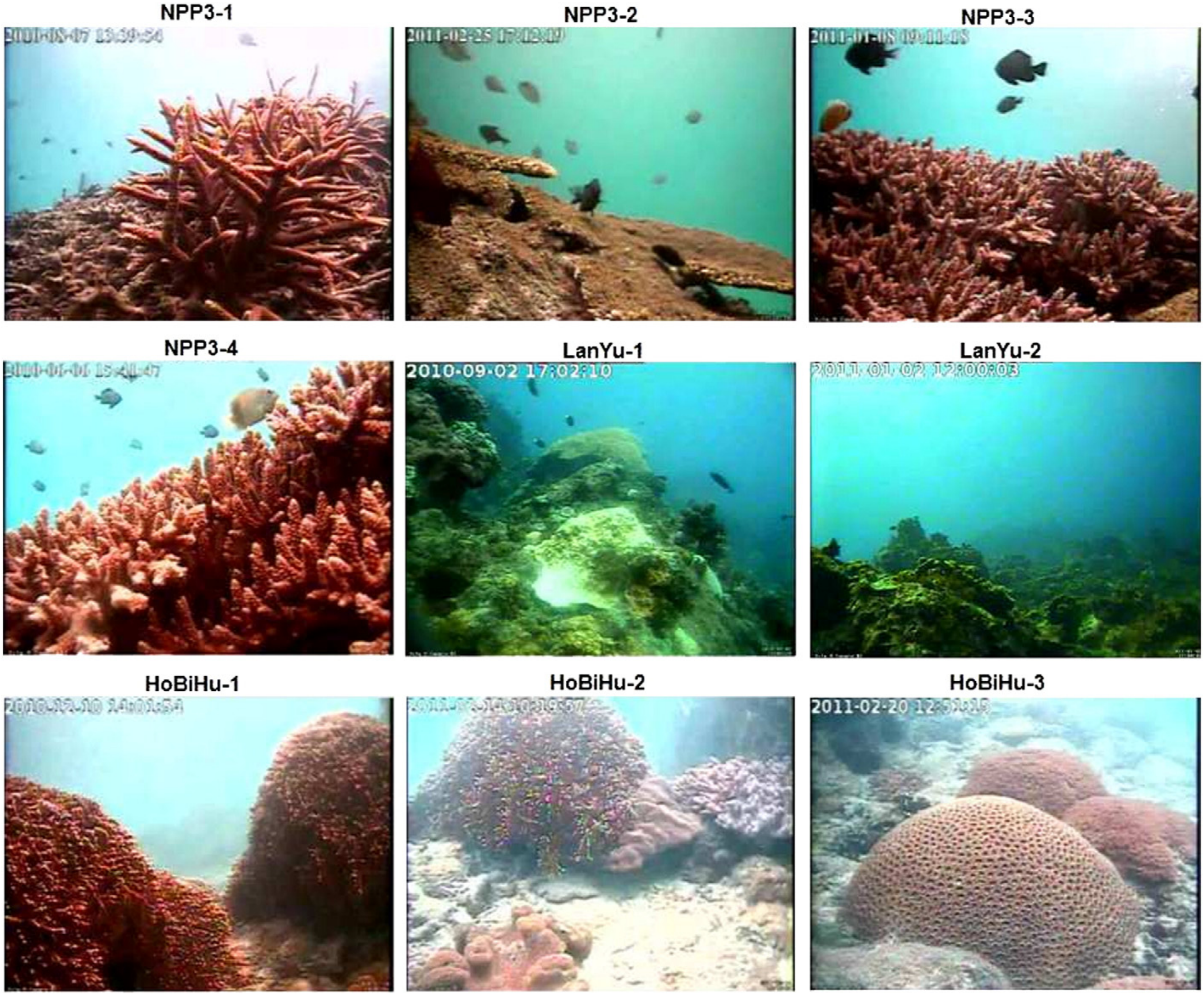


Fig. 4. Typical video footage from 9 of the cameras that record the coral reefs in southern Taiwan.

- The target environment for the Fish4Knowledge project is the Taiwanese coral reef. Beautiful though it is, it presents many traps to an object tracking algorithms: plants and algae (both moving and static), rocks, other fish, etc are all causes for occlusions which might hide a target for too long a time for the tracker to be able to recover its location from data related to many frames earlier.
- When the frame rate of the camera is relatively low (5–10 frames per second), targets may move a long way between two consecutive frames: this requires that the tracker increase the search area for a given object; however, the downside is that a larger area generally increases the risk of misassociations, in case a similar individual appears or moves inside another fish' search region.

The approach we adopted makes use of covariance-based models and is described in (Spampinato et al., 2012). The model of a fish is represented as the covariance matrix of a set of feature vectors computed for each pixel of the object. Each vector contains the pixel's (x, y) coordinates, RGB values, hue value and the mean of the grayscale histogram in a 5 × 5 window. Model similarity is performed as a comparison between covariance matrices; however, since covariance matrices do not lie in a Euclidean space (i.e. a subtraction between covariance matrices does not necessarily return a covariance matrix), we adopted Förstner's distance (Forstner and Moonen, 1999) to compare two such models:

$$\rho\left(C_i, C_j\right) = \sqrt{\sum_{k=1}^{d} ln^2 \lambda_k\left(C_i, C_j\right)} \tag{1}$$

where d is the order of the matrices and $\{\lambda_k(C_i,C_j)\}$ are the generalized eigenvalues of covariance matrices C_i and C_j, computed from

$$\lambda_k C_i x_k - C_j x_k = 0 \quad k = 1 \cdots d. \tag{2}$$

Given these premises, at each frame the following steps are performed:

- The output of the fish detection module is used to check if new fish have appeared into the scene. If so, a new *tracked fish* is added to the list of targets managed by the tracker.
- For each tracked fish, find its new location in the scene. The search window is computed based on the fish' average speed and direction up to the previous frame. Then, for each candidate region within the window, the corresponding covariance matrix is computed. The new location of the fish is set to the region which

is most similar (according to Förstner's distance) to the fish' model.

- If a tracked fish cannot be located with sufficient accuracy in the current frame, initialize a counter indicating the number of frames for which it has been missing. If the counter reaches a certain value, remove the fish from the tracker's list. To compensate for the fact that a missing fish might have moved from its original location, the search area in the following frames is increased proportionally to the missing-frame counter.

Similarly as for fish detection, we also added a tracking post-processing module for the online evaluation of the tracker's performance (described in (Spampinato and Palazzo, 2012b)). This module analyses each *tracking decision* (i.e. the association between two consecutive appearances of a fish) and estimates the likelihood that the evaluation be correct, based on appearance/geometrical variations and motion regularity. This information can be used to filter out trajectories which have a low average tracking score, or as indication for the user interface module to select only results having a minimum degree of certainty.

5.4. Species recognition

Species recognition uses the output from the fish detection and tracking module. This gives the appearance and segmentation of the fish in the videos. Based on this, the species recognition first performs feature extraction, determining interesting features to separate the different fish species. Given the features values, a hierarchical classification method is used to determine the species. Currently, the species recognition software recognizes the 15 most common species, which covers about 96% of all seen species. In the case of other species, not enough training data is available to accurately recognise these species. The species distribution in the video footage is very imbalanced which will be discussed in Section 7.1.

5.4.1. Feature selection

Pre-processing procedures are employed to improve the contour and align all fish to a similar pose. The Grabcut algorithm (Rother et al., 2004) segments fish from the background, afterward the fish orientation is computed by using curvature of the fish contour to align all fish horizontally where the head of the fish is located on the right, which is explained in (Huang et al., 2012a).

After this, 66 types of feature are extracted (Huang et al., 2012b). These features are a combination of colour, shape and texture properties in different parts of the fish such as tail/head/top/bottom, as well as the whole fish. We use a normalized colour histogram in the Red&Green channel and the Hue component in HSV colour space. These colour features are normalized to minimize the effect of illumination changes. To describe the fish texture, we calculate the co-occurrence matrix, Fourier descriptor and Gabor filter. The grey level co-occurrence matrices describe the co-occurrence frequency of two grey scale pixels at a given distance. The frequency is calculated for several orientations λ. From the co-occurrence matrices, we compute Contrast, Correlation, Energy, Entropy, Homogeneity, Variance, Inverse Difference Moment, Cluster Shade, Cluster Prominence, Max Probability, Auto correlation, Dissimilarity. These 12 features are useful as they are the first features selected by the feature selection procedure. Histograms of oriented gradients and Moment Invariants, as well as Affine Moment Invariants, are employed as the shape features. Furthermore, some specific features like tail/head area ratio, tail/body area ratio, etc. are also included. All values in the feature vector are normalized by subtracting the mean and dividing by the standard deviation (z-score normalized).

5.4.2. BGOT-based hierarchical classification

A hierarchical classification method is applied for fish recognition by using a Balance-Guaranteed Optimized Tree (BGOT) (Huang et al., 2012a). Firstly, the BGOT algorithm arranges more accurate classifications at a higher level and leaves similar classes to deeper layers. Secondly, it keeps the hierarchical tree balanced to minimize the max-depth and control error accumulation. This method controls the error accumulation in hierarchical classification and, therefore, achieves better performance.

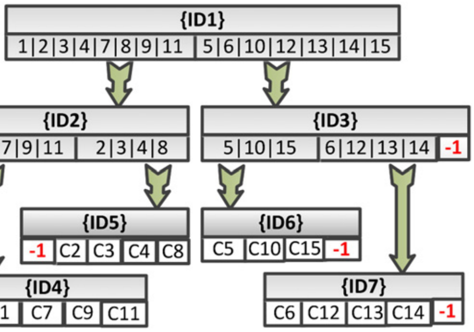


Fig. 5. A Balance-guaranteed Optimized Tree computed from training data is shown, where the leafnodes contain classifiers that either separate the fish in more subclasses or reject the fish for a particular subnode (shown by the "-1" branch), because it is not similar to the fish species in that particular node. Rejected fish are classified by a flat Support Vector Machine in this case.

As well as the BGOT tree (Huang et al., 2012a), two additions appear in this paper: node rejection and trajectory voting. The node rejection (Fig. 5) algorithm aims at controlling error accumulation. It adds a "-1" branch at each node. This branch contains all hidden classes which are classes (species) that should not be present in that node. Any fish that is classified as "-1" will be rejected, and these rejected individuals are re-classified by using a flat SVM.

Trajectory voting (Fig. 6) is used to minimize the environmental influence. As all fish are freely swimming in an environment with varying illumination and detected fish might have different orientations and appearances, so the recognition results can vary even for a fish in the same trajectory. The trajectory based voting mechanism is applied after classifying the fish in the individual frames. It combines the single classification results. The trajectory voting method enhances the fish recognition accuracy by exploiting the consistency in labels expected from tracking each fish individually. A voting mechanism (winner-take-all (Maass, 2000)) is then carried out within each group which reduces the misclassification rate.

For each detected fish, the classifier gives a certainty score, to quantify how well it is able to classify a fish based on the observed features. For a trajectory, the average certainty score of all the detected fish in the trajectory is obtained for the species determined by the majority vote.

6. User interface

6.1. Introduction

The primary goal of the interface is to support marine ecologists' research by providing means to analyse the automatically processed video data. Such analysis can be used for instance to discover ecological phenomena such as changes in fish abundance and species composition over time and area, as well as patterns in these changes.

In a preliminary interview[8] with potential users, large parts of our users' research questions concern fish counts. That is, the number of fish detected and recognized at certain time periods and in certain areas. The counts of fish are typically relevant to answer research questions concerning fish abundance, species composition and richness, growth rate of certain fish species.

[8] http://homepages.inf.ed.ac.uk/rbf/Fish4Knowledge/DELIVERABLES/Del21.pdf.

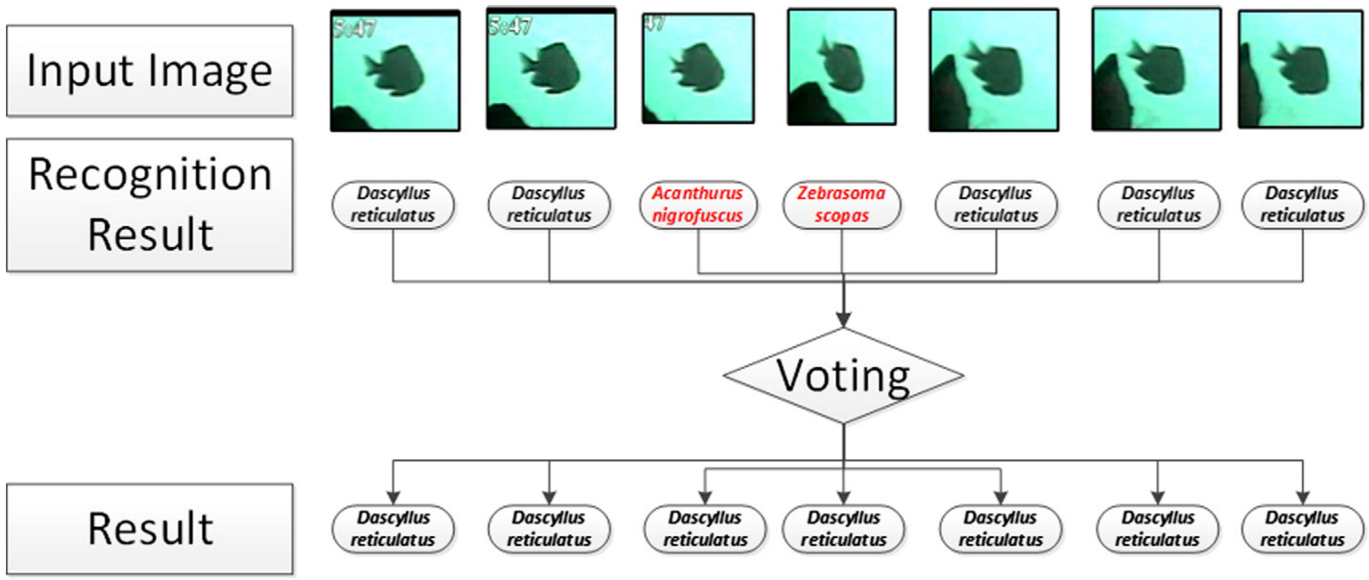


Fig. 6. An example of trajectory voting is shown where we use a winner-take-all strategy.

Counting the fish is a non-trivial task. In particular, the data collection and analysis methods within our system leads to two constraints in obtaining meaningful counting results. First, unlike divers whose vision can be wide and far, cameras have a fixed and limited viewport. Thus when fish swim in and out of the viewport of a camera, they may be counted multiple times. Second, VIP components make errors, which leads to inaccurate counting results.

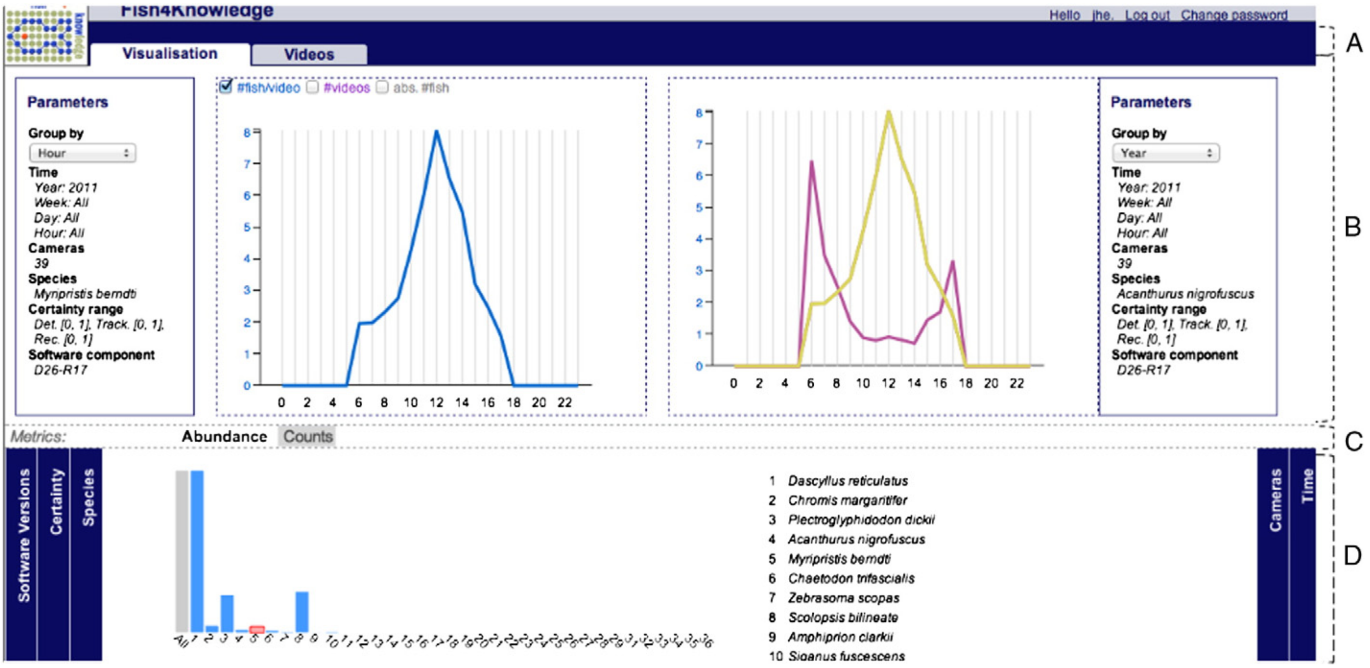
Given these constraints, as a starting point, we provide three types of quantities: the absolute counts, the counts normalized by the number of videos (e.g., relative counts), as well as the number of videos based on which the counts are generated. Notice that in our case many resident species are frequently re-observed by the cameras as well, where a direct interpretation of these counts might be difficult for marine ecologist. Linking knowledge of the fish species assemblage with observation in video (Holbrook and Schmitt, 2002) might give ecologist more understanding in the count this kind of system presents. Although further corrections of, e.g., different fish swimming rates, or VIP error rates, are needed, we leave these for future investigation.

Given the characteristics of this system and preliminary user requirements, the user interface of the first system prototype focusses on providing users with means of data exploration and validation in terms of trends in fish counts.

6.2. Interface composition

Fig. 7 shows an overview of the first prototype user interface. The interface is divided into 4 zones.

zone A. This zone contains options for exploring or validating data (i.e., result of video analysis by the VIP components). Currently

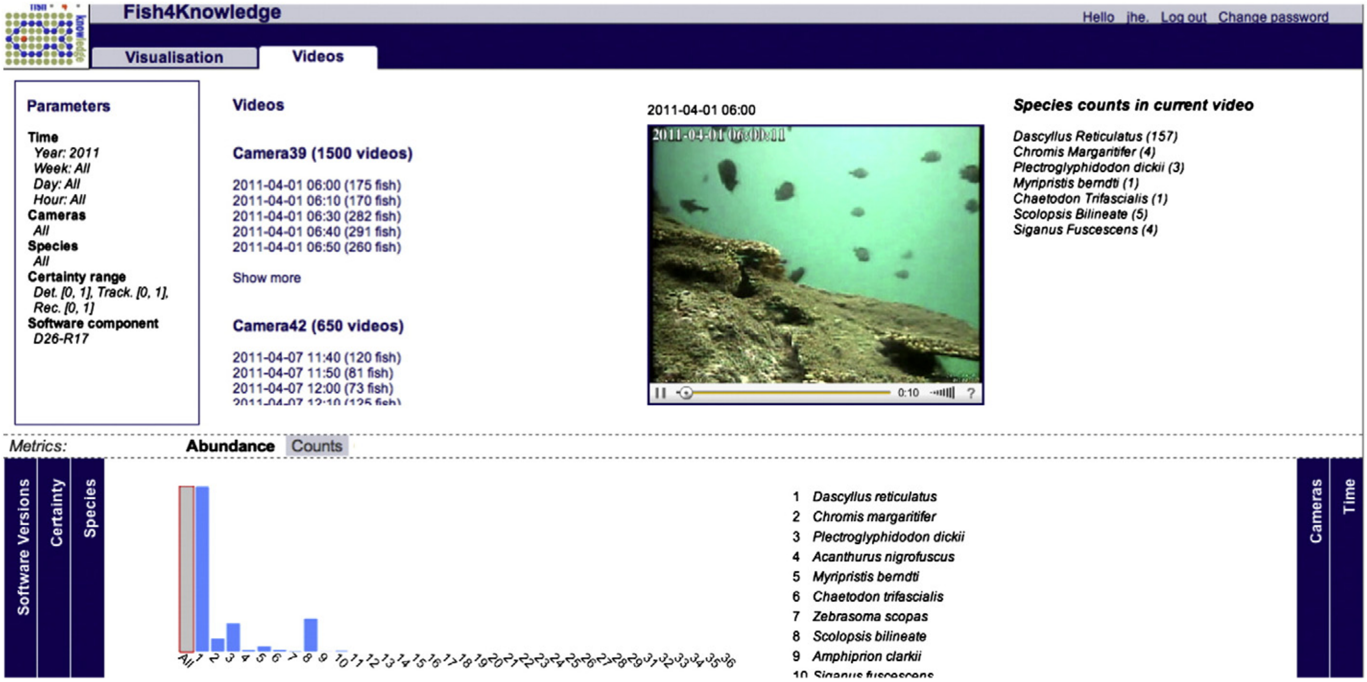

Fish4Knowledge
Hello jhe. Log out Change password
Visualisation
Videos
☑ #fish/video □ #videos □ abs. #fish
0
1
2
3
4
5
6
7
8
0 2 4 6 8 10 12 14 16 18 20 22
Metrics:
Abundance
Counts
Software Versions
Certainty
Species
Cameras
Time
All 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36
A
B
C
D

Parameters
Group by
Hour
Time
Year: 2011
Week: All
Day: All
Hour: All
Cameras
39
Species
Myripristis berndti
Certainty range
Det. [0, 1], Track. [0, 1], Rec. [0, 1]
Software component
D26-R17

Parameters
Group by
Year
Time
Year: 2011
Week: All
Day: All
Hour: All
Cameras
39
Species
Acanthurus nigrofuscus
Certainty range
Det. [0, 1], Track. [0, 1], Rec. [0, 1]
Software component
D26-R17

1 Dascyllus reticulatus
2 Chromis margaritifer
3 Plectroglyphidodon dickii
4 Acanthurus nigrofuscus
5 Myripristis berndti
6 Chaetodon trifascialis
7 Zebrasoma scopas
8 Scolopsis bilineata
9 Amphiprion clarkii
10 Siganus fuscescens

Fig. 7. An overview of the user interface, in the “visualization” view.



Parameters

Time
Year: 2011
Week: All
Day: All
Hour: All
Cameras
All
Species
All
Certainty range
Det. [0, 1], Track. [0, 1], Rec. [0, 1]
Software component
D26-R17

Videos

Camera39 (1500 videos)

2011-04-01 06:00 (175 fish)
2011-04-01 06:10 (170 fish)
2011-04-01 06:30 (282 fish)
2011-04-01 06:40 (291 fish)
2011-04-01 06:50 (260 fish)

Show more

Camera42 (650 videos)

2011-04-07 11:40 (120 fish)
2011-04-07 11:50 (81 fish)
2011-04-07 12:00 (73 fish)
2011-04-07 12:10 (125 fish)

2011-04-01 06:00

2011-04-01 06:00:11

Species counts in current video

Dascyllus Reticulatus (157)
Chromis Margaritifer (4)
Plectroglyphidodon dickii (3)
Myripristis berndti (1)
Chaetodon Trifascialis (1)
Scolopsis Bilineate (5)
Siganus Fuscescens (4)

1 *Dascyllus reticulatus*
2 *Chromis margaritifer*
3 *Plectroglyphidodon dickii*
4 *Acanthurus nigrofuscus*
5 *Myripristis berndti*
6 *Chaetodon trifascialis*
7 *Zebrasoma scopas*
8 *Scolopsis bilineate*
9 *Amphiprion clarkii*
10 *Siganus fuscescens*

Fig. 8. Data validation with videos.

two options are provided: "Visualization" and "Video". Option "Visualization" displays graphical visualizations of chosen metrics over selected data. Option "Videos" displays videos from which the analysis data is extracted so that users can validate their observations by watching these videos.

zone B. It is the area where the selected visualization or videos are displayed.

zone C. This zone provides a set of *metrics* that can be used to characterize ecological phenomena. At the moment we focus on the fish abundance (i.e., counts of fish).

zone D. This zone contains a set of *filters*. Each filter represents a perspective of the data, including perspectives concerning the video data collection process such as time, location, fish species, as well as perspectives concerning data generating process such as certainty of VIP components in their outputs and the software version that has been used to process the videos.

On the one hand, users can use these filters to select data of interest, while on the other hand, the filters provide a summary of the selected data from a given perspective of the data.

Below we describe the above components with more details.

6.2.1. Metrics

The metrics we consider within the system are derived from the preliminary interviews with marine biologists/ecologists as mentioned above. At the moment, we have implemented the metric for fish abundance.

6.2.2. Filters

Filters provide users with an overview on different facets of the data generated by the computer vision components. A facet refers to a certain category of the values within the data that can be used to filter the data, e.g., time and fish species. The following filters are created:

Time. Our data covers the fish detection, tracking and recognition results extracted from video footages over a 3-year period. With the time filters users have the overview of the fish counts summarized in terms of year, week, day and hour.

Cameras. Our video data are recorded with 9 underwater cameras in 3 areas in the Taiwanese sea. Camera filters summarize fish counts with respect to data obtained from different cameras.

Species. Species filters summarize the fish counts with respect to different fish species. At the moment, the fish recognition components are able to recognise the 15 most frequently encountered species.

Certainty. The results of fish detection, tracking and recognition all come with a certainty score, indicating the confidence of the algorithm in the correctness of its prediction.[9] Certainty filters summarize the fish counts over different ranges of certainty scores.

Software components. The computer vision software modules have been constantly improved and multiple software modules can process the data to verify trends. Software modules filters summarize the fish counts for the different software modules.

The filters use bar charts to characterize the distribution of fish counts in a facet. By selecting these bars, users select the data of interest. This data is then visualized or displayed as relevant videos in zone B, as well as in the rest of the filters. Notice that, this way, the filters are not independent from each other: that is, a selection in one filter determines the data visualized by the rest of the filters.

With these dependencies, users are able to have a tangible feeling of how different factors influence their observations and how these factors interact with each other.

For instance, the distribution of fish counts over different species recorded by camera X may be different from that recorded by camera Y, as the cameras are located in different areas.

6.2.3. Data visualization

The selected data and metric can be visualized in two views: (1) aggregated view (left); and (2) comparative view (right), as shown in Fig. 7. For a selected set of data, the aggregated view shows the temporal

[9] In term of fish detection, correctness refers to whether or not a fish is correctly identified as a fish and non-fish object is identified as "not-a-fish"; in terms of fish tracking, correctness refers to whether a sequence of fish instances belonging to the same fish is correctly identified as a single fish; and in terms of fish recognition, correctness refers to whether a fish species is correctly identified.

Table 2
Description of the videos used as ground truth.

Video	Frame rate (fps)	Frame size	Description	Number of objects
1	5	320 × 240	Normal conditions	1058
2	5	320 × 240	Normal conditions	1656
3	5	320 × 240	Murky water	1284
4	5	320 × 240	Murky water Fish crowds	5477
5	5	320 × 240	Sun gleaming on surface Fast lighting variations	3072
6	5	320240	Fish crowds Algae on lens	16321
7	5	320 × 240	Dynamic background Encoding problems	1927

change of the fish counts. That is, the total fish counts over the selected data grouped by certain time unit, e.g., year, week, and hour. The parameters used to select the data (i.e., selection over the filters) are shown on the left side of the aggregated view. After constructing a curve in the aggregated view, users can click on it and put it in the comparative view. Thus multiple curves (i.e., quantities, trends) can be compared under the same scale. When the mouse hovers over a line in the comparative view, the corresponding parameter setting of the line shows in the right side of the view.

6.2.4. Data validation with videos

Users can check the content of videos to validate the observation they obtain on the selected data in the "Videos" tab, see Fig. 8. Same as in the "Visualization" tab, parameters used to selected the data is displayed on the left side of the screen. Videos are listed together with the total number of fish detected from each of them. By clicking on a video link, users can watch it playing on the right side of the screen. Next to the playing video, we list the running count of each species.

7. Evaluation of video and image processing

7.1. Data to evaluate the VIP software modules

In the previous section, we showed how marine ecologists can verify part of the data by looking at the actual count of fish in the videos. This is however limited to certain subsets of the data, because we cannot expect marine ecologists to verify the current amount of imagery equating to forty thousand videos, each of 10 min duration. Although verification of this data is impossible, evaluation on a subset of this data is possible (where we assume this subset is representative of all videos).

The first requirement for the evaluation of most computer vision algorithms is the existence of a ground truth which the algorithms' results can be compared to. In the case of fish detection, we chose a subset of 11 videos from the whole footage repository and performed a manual labelling of each individual frame, identifying all fish present in the scene, drawing the contours and specifying the associations between appearances of the same fish in different frames. All videos are 10 min long and have varying resolutions and frame rate, and were chosen to cover as many different scene conditions (Murky water, Algae on lens, Encoding Problems) as possible. Table 2 describes each of these videos, specifying its technical information and the scene features it presents.

Evaluation of the fish recognition software is performed on a set of fish images extracted from videos where the species identification is performed by human annotators. Part of fish images are from the

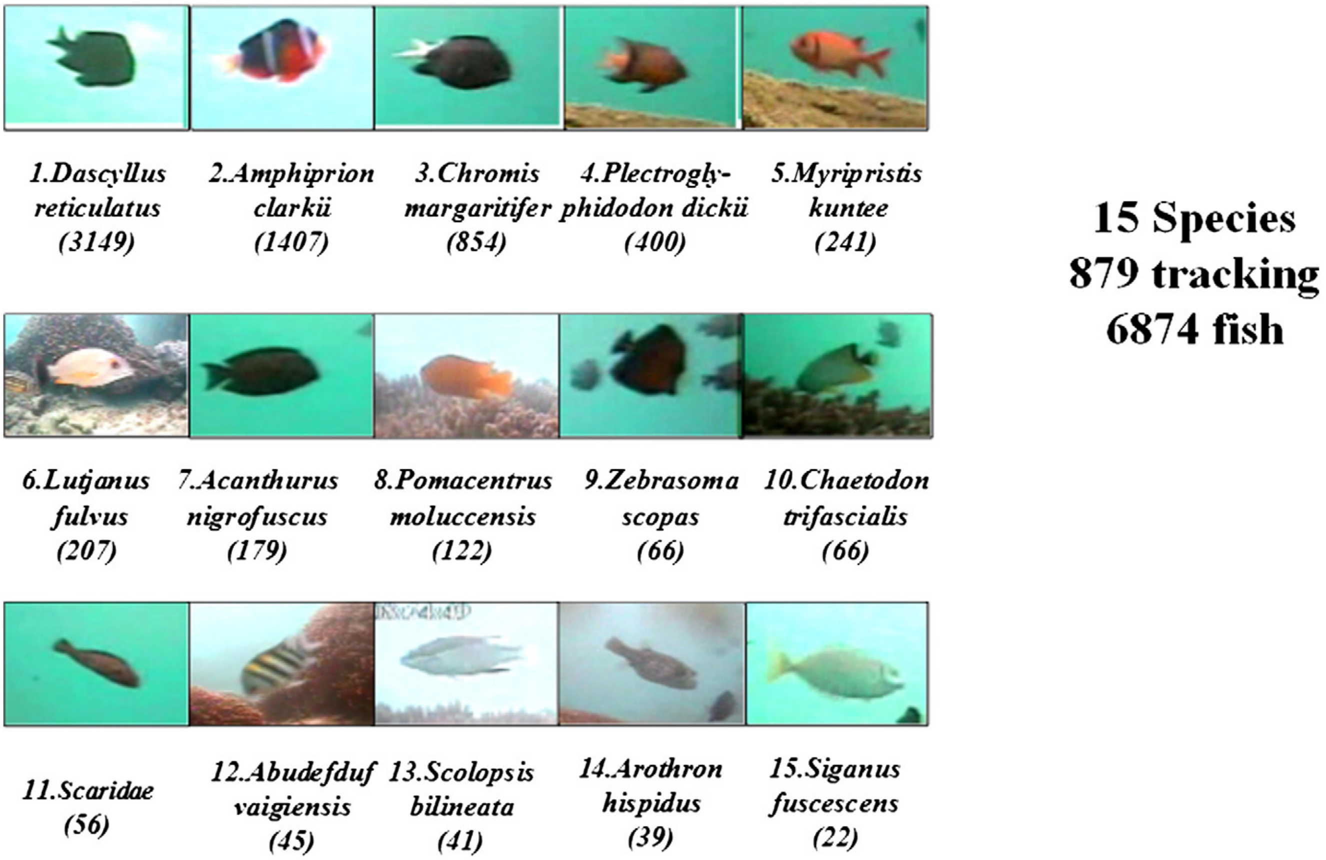


Fig. 9. Top 15 species of fish in underwater videos.

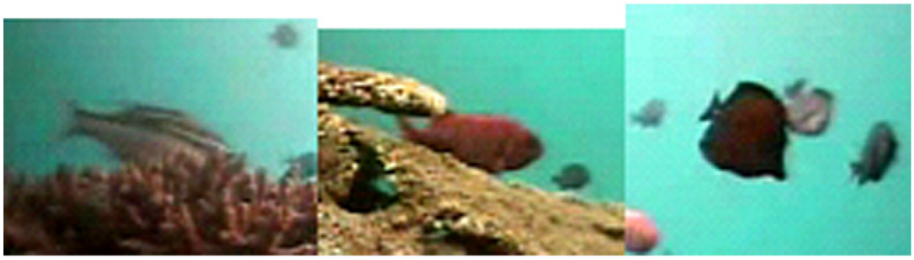


Fig. 10. Difficult examples for species recognition, in the first two cases the fish are occluded by objects where the automatic system does not have correct contour information to classify the fish, while in case of fish overlapping from the camera view point, it is difficult to determine a clear border where the software can avoid classifying two fish as a single fish.

same dataset as is used for the fish detection. However, the other part of the dataset is obtained using automatically detected fish in the videos to become more invariant against small segmentation errors in the detection process. Both experts and non-experts identified fish images by grouping them together based on their appearance, where we use clustering to speed up this process and combine the experts' and non-experts' identification in a smart manner (Boom et al., 2012). This allowed us to obtain 6874 fish images of the 15 different species shown in Fig. 9. This figure shows the fish species name and the numbers of detections. The data is very imbalanced where the most frequent species is about 150 times more common than the least one. Many fish images are low quality: blurred, occluded by other fish or background objects, which include coral, the sea flower and open sea. Fig. 10 shows some difficult examples for species recognition. The fish species are manually labelled by following instructions from marine ecologists.

7.2. Fish detection results

The results are shown in terms of *detection rate (DR)* and *false alarm rate (FAR)*, defined as:

$$DR = \frac{TP}{TP + FN} \quad (3)$$

$$FAR = \frac{TP}{TP + FP} \quad (4)$$

where *TP*, *FP* and *FN* represent respectively the number of true positives, false positives and false negatives.

The detection evaluation of the fish detection algorithms was performed at *blob level*, in order to verify whether an algorithm is able to assess the presence or absence of a fish in a frame, and at *pixel level*, in order to check how close the extracted contour is to the actual contour (in this case, the relevant evaluation quantities are *PDR* and *PFAR*, defined similarly as *DR* and *FAR*, but related to whether a given pixel belongs to the object or not). For both levels, the results are shown with the application of the detection post-processing filtering module. Table 3 shows the performance of the set of algorithms that we tested in the underwater domain. The corresponding ROC curve for the selected algorithms is also shown in Fig. 11 (the curve for APMM is not reported because of the high processing times to compute the values of the curve). From Fig. 11, we observe that ViBe has the best performance detection rate (which is the reason that this software module is used to process most of the data), while Adaboost is able to filter out more incorrect detections because of the low false alarm rate allowing marine ecologists to compare the different results.

Table 3
Evaluation of the motion detection algorithms in the underwater domain.

	DR	*FAR*	*PDR*	*PFAR*
Adaboost	72.3%	11.4%	91.2%	18.0%
GMM	69.6%	13.2%	90.6%	19.2%
APMM	70.4%	19.2%	83.2%	17.2%
IM	72.9%	17.4%	87.8%	20.1%
ViBe	79.9%	18.3%	93.2%	12.6%

7.3. Fish tracking results

Fish tracking evaluation assesses the ability of the tracker to follow a fish. However, no universally accepted comparison strategy exists for object trajectories. (Porikli, 2004) explains the reasons why a comparison of two trajectories represented as sequences of points is not reliable. For this reason, we adopted three quantities which give indications on the goodness of a tracking algorithm from different points of view:

- *Correct counting ratio (CCR)*: percentage of correctly identified fish out of the total number of ground truth trajectories; we say that two trajectories match if they have more than 50% of fish detections in common.
- *Average trajectory match (ATM)*: average percentage of common fish detections between a ground truth trajectory and the corresponding best-matching tracker trajectory.
- *Correct decision ratio (CDR)*: we define a *tracking decision* as an association between two objects in two consecutive frames; *CDR* is the percentage of correct associations, and provides further information on how good the tracker is in following an object, which is not necessarily implied by a high *ATM* value, as shown in Fig. 12.

Table 4 shows these evaluation scores for the covariance-based tracking that we used in this work, and for the CAMSHIFT (Bradski, 1998) algorithm used in (Spampinato et al., 2008). It can be seen how the covariance-based tracker is able to obtain the best performances when using either of the three scores.

7.4. Species recognition results

The experiment is based on the 6874 fish images, where a 5-fold cross validation procedure is used in our experiments. Fish images from the same trajectory sequence are either only in the training or in the test set. Sequential forward feature selection is applied at each tree node. The recognition results are listed in Table 5 for three performance metrics.

Three performance metrics are employed to evaluate the accuracy of the proposed system. The first metric is Average Recall (AR) over all species. It describes on average how many fish are correctly recognized for each species. This score is more important to our experiment because of the imbalance in the species. Generally, given True Positive/False Positive/False Negative, the AR is defined as:

$$AR = \frac{1}{c}\sum_{j=1}^{c}\left(\frac{TruePositive_j}{TruePositive_j + FalseNegative_j}\right) \quad (5)$$

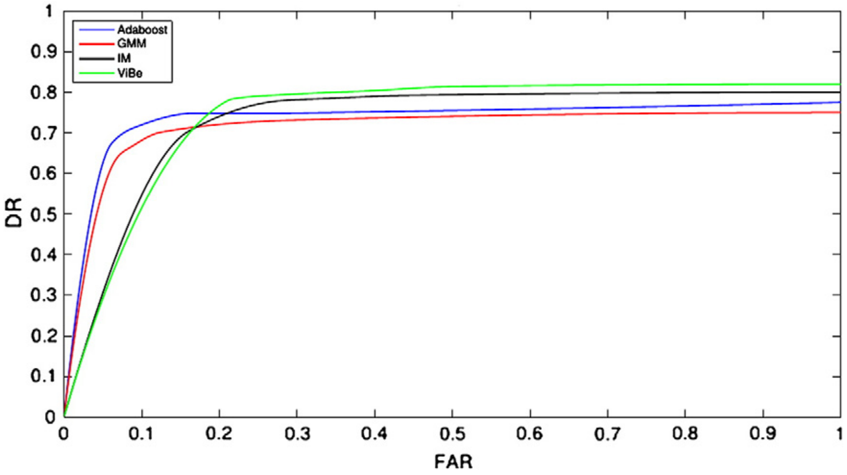


Fig. 11. ROC curve for the motion detection algorithms used in this work.

where c is the number of species. The second score is Average Precision (AP) over all species. It is the probability that the classification results are relevant to specified species, as shown below:

$$AP = \frac{1}{c}\sum_{j=1}^{c}\left(\frac{TruePositive_j}{TruePositive_j + FalsePositive_j}\right) \quad (6)$$

The third metric is the accuracy over all samples (Accuracy over Count, AC), which is defined as the proportion of correct classified each fish in the whole dataset. The AC is calculated as following:

$$AC = \frac{\sum_{j=1}^{c} TruePositive_j}{\sum_{j=1}^{c}\left(TruePositive_j + FalsePositive_j\right)} \quad (7)$$

We compare the hierarchical classification against the flat SVM classifier (AR = 76.71%) and taxonomy based hierarchical tree. The taxonomy methodology is based on the synapomorphies (common characters of fish) and indicates the distinction between species which helps organize similar species for later processing. As a result, the taxonomy based hierarchical tree achieves a higher AR (77.16%) than the flat SVM but is worse than the automatically generated hierarchical tree (85.75%), which chooses the best splitting by exhaustively searching all possible combinations while remaining balanced. The search procedure takes several hours and a possible improvement is to integrate the hierarchical method with domain knowledge like taxonomy, which helps organize similar species for later processing, instead of exhaustive searching.

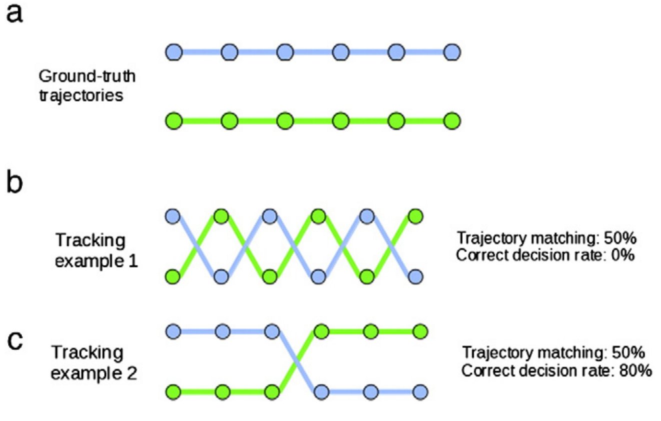


Fig. 12. Fig. 12(a) shows two ground truth trajectories of two fish, whereas the other two images represent two examples of tracking output. In Fig. 12(b) although the tracker fails at each tracking decision the average trajectory matching score is 50%, whereas the correct decision rate is 0%. In Fig. 12(c) the tracker fails only in one step and the average trajectory matching score is 50% (as the previous case) whereas the correct decision rate is 80% (4 correct associations out of 5).

We also have compared with the classifier in (Huang et al., 2012a), which was the first software for species recognition in our system. We use the same dataset, and choose the top 10 species because the previous component can only recognise these 10 species. The AR score shows that our new component achieves 81.0%, which is 14% higher than the old component (65.6%). Better species recognition modules can be easily included later and will given more information to marine ecologists.

8. Validation of system observations

As described in Section 6, our user interface is characterized by two features: data exploration and validation. Here, without losing generality, we describe a scenario to illustrate how the data exploration and validation functionalities of the current UI can help users to discover and understand the patterns present in the data.

Although our system can show new and interesting patterns in fish counts (see Section 3), ecologists have to be aware of the risk of overrelying on automatic systems (in which presented data may contain various random and systematic errors). The ability of the system to validate observations is necessary to determine which observations are biological meaningful and which are related to possible systematic errors. For instance, while the observation of a sudden drop in the number of fish detected by our system may reflect the decreasing of the actual

Table 4
Description of the videos used as ground truth.

	Covariance tracker	CAMSHIFT
CCR	91.3%	83.0%
ATM	95.0%	88.2%
CDR	96.7%	91.7%

Table 5
Fish recognition results. Our proposed result is in Bold font. We add the standard deviation of AR/AP/AC over 5 fold validation. * means the AC is a significant improvement over other methods at 95% confidence level.

Method	AR (%)	AP (%)	AC (%)
(lr)2-4 SVM (fs)	76.71 ± 5.93	81.47 ± 5.27	93.50
taxonomy	77.16 ± 6.29	80.80 ± 6.92	93.76
BGOT method	**85.75** ± **5.64**	**91.30** ± **8.73**	* **97.21**

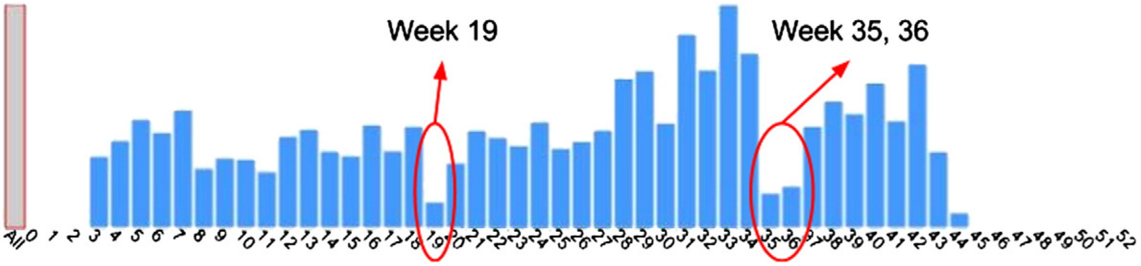
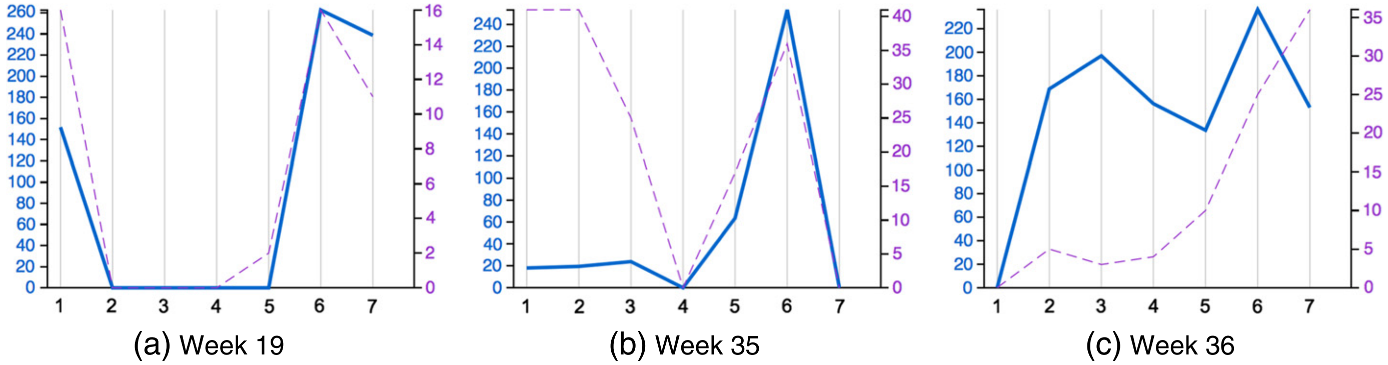


Fig. 13. Fish counts in 2011, grouped by week of the year.



(a) Week 19 (b) Week 35 (c) Week 36

Fig. 14. Fish counts in the three weeks 19, 35 and 36, grouped by day (1–7). Dashed curves are counts of videos, and solid curves are the average count of fish per video.

fish abundance, which is ecologically interesting for further analysis, it may also be caused by a few corrupted videos, which is ecologically irrelevant.

Our tool is designed to assist users in quickly determining such systematic errors caused by ecologically non-relevant factors, which is otherwise impossible to detect given the large number of videos (30,048 videos). Below, we show such a case.

8.1. Observation validation scenario

In this scenario, we discuss how our interface can be used to explore causes of variance in fish counts. Notice that here, we are *not* interested in the variance of fish counts due to randomness, i.e., the variance of fish counts as a random variable. Rather, we are interested in finding out "external" factors that influence the observed fish counts.

Fig. 13 shows the per-week fish counts in 2011 processed by the ViBe method for fish detection and the 15 species BGOT method for species recognition.[10]

We notice that among the processed video data (videos between week 0–2 and 44–52 have not finished processing), weeks 19, 35 and 36 have exceptionally low counts compared to the rest weeks. Why do these weeks have low numbers of fish detected/recognized? Is it caused by the video processing system? or is it caused by something happening in the natural environment?

The videos in our system are not processed sequentially, e.g., due to parallel processing and situations where processing failed (due to system crashes, unavailable videos, etc). Our first suspect would be that on these weeks, lower numbers of videos have been processed. While users may not know this fact of our system, they can observe this phenomenon from our visualizations. Fig. 14 shows the per-day fish counts for weeks 19, 35, and 36, along with the number of processed videos. We see that for week 19 and 36, indeed there are days where the number of processed videos is low: day 2, 3, 4 and 5 of week 19, and day 1, 2, 3, 4, and 5 of week 36. In addition, with a close look at the Y-axes of these plots, we see that the numbers of videos processed in week 19 are generally lower than weeks 35 and 36. This observation suggests that it is very likely that the low fish counts in week 19 and 36 are due to the low number of videos processed.

What surprises us in Fig. 14 is that, in week 35 on day 1, 2, and 3, while the number of videos processed are not low, the average per-video fish counts are low. That is, the low fish counts of week 35 *cannot* be explained by the lack of processed videos. By looking at the videos of these days (Fig. 15), we immediate see the reason why there are very few fish detected: the videos recorded on these days show murky water and unclear images.

Moreover, it is not just one camera, but *all* cameras have the same conditions, which suggests that it is not due to the deficiency of a single camera.

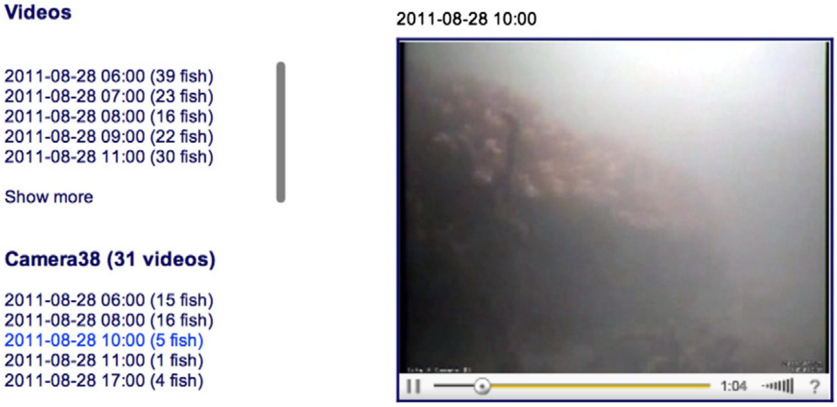
With a little extra research, it is not difficult to discover that the first 3 days of week 35, 2011 is in a typhoon period: typhoon Nanmadol passes Taiwan between 26–28th August, 2011,

In summary, this scenario shows how our current system can be used to explore and dig into available data for explanations of interesting observations.

9. Discussion

A new data-driven methodology for marine ecology to observe long-term and continuous trends in local fish assemblages has been described in this paper. In this case, underwater videos are analysed by automatic computer vision (Huang et al., 2012a; Spampinato et al., 2010) software that provides the data which afterwards can be analysed by marine ecologists. In comparison with other large scale data collection projects in ecology like ebird (Sullivan et al., 2009) and flora observations (Auer et al., 2011; McGuire et al., 2008), this project is different because data collection is not performed by volunteers and our data cannot be represented geographically. A similarity between especially the ebird (Sullivan et al., 2009) project is that with the data also comes uncertainty. Dealing with this uncertainty is the real challenge

[10] At the time when this paper was written, we used the version of the detection, tracking and recognition that had processed most of available videos in 2011.



Videos

2011-08-28 06:00 (39 fish)
2011-08-28 07:00 (23 fish)
2011-08-28 08:00 (16 fish)
2011-08-28 09:00 (22 fish)
2011-08-28 11:00 (30 fish)

Show more

Camera38 (31 videos)

2011-08-28 06:00 (15 fish)
2011-08-28 08:00 (16 fish)
2011-08-28 10:00 (5 fish)
2011-08-28 11:00 (1 fish)
2011-08-28 17:00 (4 fish)

2011-08-28 10:00

Fig. 15. Videos of week 35.

in this kind of project. Efforts by the ebirds project, where they model experts versus novices (Yu et al., 2010) performance should also be explored in this project to model experts versus automatic software performance.

Currently, the automatic computer vision (Huang et al., 2012a; Spampinato et al., 2010) is performed by a couple of algorithms developed by the same group of people. However, the system is capable of running other software for fish detection/recognition given that it saves the results according to our database definitions. It would be interesting to compare the fish recognition methodologies (Lee, 2004; Rova et al., 2007; Toh et al., 2009) with eachother. Also if new software is created, it would be interesting to see how the performance differs from the existing software. Maintenance of software by the marine ecology community is essential, where plankton classification software (Benfield et al., 2007) can be seen as a positive example. Maintenance of data is also essential, because modern recognition software often learns from large datasets of images, where new challenges in domain specific image recognition (Khosla et al., 2011) can bring better methodologies. Resources like the fishbase.org can try to collect and support these kind of challenges, while standardization in data and software is necessary (Antoniou and Harmelen, 2004) for comparison of domain specific recognition software.

The data obtained in the system should be studied further by marine ecologists. The user interface is open for everyone at http://f4k.project.cwi.nl/data1/ui/.[11] Relating the data obtained by our user interface to earlier observation and assemblage measures performed in (Jan et al., 2007), which describes the original camera monitoring side could be done. Changes with respect to the recording and camera views could be taken into account. Relating results extracted by our system to other paper in marine ecology will also be interesting. For example, the research by (Holbrook and Schmitt, 2002) on *Dascyllus flavicaudus* and *Dascyllus trimaculatus* might to be related to the dominant fish *Dascyllus recticulatus* in our data and similar trends might be observed, while new long-term observations might be discovered.

The research tool we have presented gives marine ecologist a user interface to analyse long-term and continuous video content using automatic video recognition software. This is the first prototype of such a complex system and has by no means the status of a production system. We are aware of the limitations of the current system, as well as the limitations of the data obtained by video recording methods, cf. Table 1. However, it is important that the marine ecology community is aware of the availability of this type of data and of tools that can process and mine such data. With a larger user base and the assistance from the marine ecology community, we can extend and improve this tool to fit into specific needs of the users and generate more valuable results for the marine ecology community.

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). We would like to thank the reviewers for their suggestion and Kwang-Tsao Shao of Biodiversity Research Center in Taiwan as well as the all the project members: Emmanuelle Beauxis-Aussalet, Sun-In Lin, Gayathri Nadarajan, Yun-Heh Chen-Burger, Jacco van Ossenbruggen, Daniela Giordano, Lynda Hardman for their contribution in this project.

References

Antoniou, G., Harmelen, F., 2004. Web ontology language: Owl. In: Staab, S., Studer, R. (Eds.), Handbook on Ontologies. International Handbooks on Information Systems. Springer, Berlin Heidelberg, pp. 67–92 (URL http://dx.doi.org/10.1007/978-3-540-24750-0_4).

Auer, T., MacEachren, A.M., McCabe, C., Pezanowski, S., Stryker, M., 2011. Herbariaviz: A web-based clientÃ¢â,¬â€œserver interface for mapping and exploring flora observation data. Ecol. Inform. 6 (2), 93–110.

Barnich, O., Van Droogenbroeck, M., Jun. 2011. ViBe: a universal background subtraction algorithm for video sequences. IEEE Trans. Image Process. 20 (6), 1709–1724.

Benfield, M., Grosjean, P., Culverhouse, P., Irigoien, X., Sieracki, M., Lopez-Urrutia, A., Dam, H., Hu, Q., Davis, C., Hansen, A., et al., 2007. Rapid: Research on Automated Plankton Identification.

Boom, B.J., Huang, P.X., He, J., Fisher, R.B., 2012. Supporting Ground-Truth Annotation of Image Datasets Using Clustering. ICPR.

Bradski, G.R., 1998. Computer Vision Face Tracking For Use in a Perceptual User Interface.

Cappo, M., Harvey, E., Shortis, M., 2006. Counting and measuring fish with baited video techniques—an overview. AFSB conferenc and workshop cutting-edge technologies in fish and fisheries science.

Davis, C., Thwaites, F., Gallager, S., Hu, Q., 2005. A three-axis fast-tow digital video plankton recorder for rapid surveys of plankton taxa and hydrography. Limnol. Oceanogr. Methods 3, 59–74.

Dorman, S.R., Harvey, E.S., Newman, S.J., 07 2012. Bait effects in sampling coral reef fish assemblages with stereo-bruvs. PLoS ONE 7 (7), e41538.

Ebner, B., Clear, R., Godschalx, S., Beitzel, M., 2009. In-stream behaviour of threatened fishes and their food organisms based on remote video monitoring. Aquat. Ecol. 43 (2), 569–576.

Faro, A., Giordano, D., Spampinato, C., Dec. 2011. Adaptive background modeling integrated with luminosity sensors and occlusion processing for reliable vehicle detection. IEEE Trans. Intell. Transp. Syst. 12 (4), 1398–1412.

Forstner, W., Moonen, B., 1999. A metric for covariance matrices. Tech. rep., Dept. of Geodesy and Geoinformatics. Stuttgart University.

Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55 (1), 119–139.

Hill, J., Wilkinson, C., 2004. Methods for Ecological Monitoring of Coral Reefs. Australian Institute of Marine Science, Townsville 117.

Holbrook, S.J., Schmitt, R.J., 2002. Competition for shelter space causes density-dependent predation mortality in damselfishes. Ecology 83 (10), 2855–2868 (URL http://www.jstor.org/stable/3072021).

Huang, P., Boom, B., Fisher, R., 2012a. Underwater live fish recognition using a balance-guaranteed optimized tree. Asian Conference on Computer Vision.

Huang, P.X., Boom, B.J., Fisher, R.B., 2012b. Hierarchical classification for live fish recognition. BMVC student workshop paper.

Jan, R.-Q., Shao, Y.-T., Lin, F.-P., Fan, T.-Y., Tu, Y.-Y., Tsai, H.-S., Shao, K.-T., 2007. An underwater camera system for real-time coral reef fish monitoring. Raffles Bull. Zool. 14, 273–279.

[11] We are improving this interface. The current version used in the paper is mentioned in the text. Newer version are made available at http://f4k.project.cwi.nl/demo/ui/.

Kelling, S., Hochachka, W., Fink, D., Riedewald, M., Caruana, R., Ballard, G., Hooker, G., 2009. Data-intensive science: a new paradigm for biodiversity studies. Bioscience 59 (7), 613–620.

Khosla, A., Jayadevaprakash, N., Yao, B., Fei-Fei, L., 2011. Novel dataset for fine-grained image categorization. First Workshop on Fine-Grained Visual Categorization. CVPR.

Larsen, R., Ólafsdóttir, H., Ersbøll, B., 2009. Shape and texture based classification of fish species. Proceedings of the Scandinavian Conference on Image Analysis, pp. 745–749.

Lee, D.J., 2004. Contour matching for a fish recognition and migration-monitoring system, vol. 5606. Proceedings of SPIE, Philadelphia, PA, USA, pp. 37–48.

Luo, T., Kramer, K., Goldgof, D., Hall, L., Samson, S., Remsen, A., Hopkins, T., 2004. Active learning to recognize multiple types of plankton. Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, vol. 3. IEEE, pp. 478–481.

Maass, W., 2000. On the computational power of winner-take-all. Neural Comput. 12 (11), 2519–2535. http://dx.doi.org/10.1162/089976600300014827 (Nov.).

McGuire, M., Gangopadhyay, A., Komlodi, A., Swan, C., 2008. A user-centered design for a spatial data warehouse for data exploration in environmental research. Ecol. Inform. 3 (4), 273–285.

Pattengill-Semmens, C.V., Semmens, B.X., 2003. Conservation and management applications of the reef volunteer fish monitoring program. Environ. Monit. Assess. 81, 43–50.

Pelletier, D., Leleu, K., Mallet, D., Mou-Tham, G., Hervé, G., Boureau, M., Guilpart, N., 2012. Remote high-definition rotating video enables fast spatial survey of marine underwater macrofauna and habitats. PLoS ONE 7 (2).

Porikli, F., 2004. Trajectory distance metric using hidden Markov model based representation. European Conference on Computer Vision. IEEE.

Porikli, F., 2005. Multiplicative background-foreground estimation under uncontrolled illumination using intrinsic images. Proceedings of the IEEE Workshop on Motion and Video Computing, vol. 2, pp. 20–27.

Rother, C., Kolmogorov, V., Blake, A., 2004. GrabCut: interactive foreground extraction using iterated graph cuts. ACM Trans. Graph. 309–314.

Rova, A., Mori, G., Dill, L.M., 2007. One fish, two fish, butterfish, trumpeter: Recognizing fish in underwater video. IAPR Conference on Machine Vision Applications, pp. 404–407.

Ruff, B.P., Marchant, J.A., Frost, A.R., 1995. Fish sizing and monitoring using a stereo image analysis system applied to fish farming. Aquac. Eng. 14 (2), 155–173.

Shortis, M., Harvey, E., Abdo, D., et al., 2009. A review of underwater stereo-image measurement for marine biology and ecology applications. Oceanogr. Mar. Biol. 47, 257.

Spampinato, C., Palazzo, S., 2012a. Enhancing object detection performance by integrating motion objectness and perceptual organization. Proceedings of the 21st International Conference on Pattern Recognition, ICPR, pp. 3640–3643.

Spampinato, C., Palazzo, S., 2012b. Evaluation of tracking algorithm performance without ground-truth data. IEEE International Conference on Image Processing, to appear.

Spampinato, C., Chen-Burger, Y.H., Nadarajan, G., Fisher, R.B., 2008. Detecting, tracking and counting fish in low quality unconstrained underwater videos. 3rd International Conference on Computer Vision Theory and Applications. VISAPP, pp. 514–519.

Spampinato, C., Giordano, D., Salvo, R.D., Chen-Burger, Y.H., Fisher, R.B., Nadarajan, G., 2010. Automatic fish classification for underwater species behavior understanding. Proceedings of the first ACM international workshop on analysis and retrieval of tracked events and motion in imagery streams. ACM, New York, NY, USA, pp. 45–50.

Spampinato, C., Palazzo, S., Giordano, D., Lin, F.P., Lin, Y.T., 2012. Covariance-based fish tracking in real-life underwater environment. Proceedings of the International Conference on Computer Vision Theory and Applications.

Stauffer, C., Grimson, W.E.L., 1999. Adaptive background mixture models for real-time tracking. Computer Vision and Pattern Recognition. IEEE Computer Society Conference on, 2, pp. 246–252.

Strachan, N.J.C., 1993a. Length measurement of fish by computer vision. Comput. Electron. Agric. 8 (2), 93–104.

Strachan, N.J.C., Jan. 1993b. Recognition of fish species by colour and shape. Image Vis. Comput. 11, 2–10 (ACM ID: 156583).

Strachan, N.J.C., Nesvadba, P., Allen, A.R., 1990. Fish species recognition by shape analysis of images. Pattern Recogn. 23 (5), 539–544.

Sullivan, B., Wood, C., Iliff, M., Bonney, R., Fink, D., Kelling, S., 2009. ebird: A citizen-based bird observation network in the biological sciences. Biol. Conserv. 142 (10), 2282–2292.

Toh, Y.H., Ng, T.M., Liew, B.K., 2009. Automated fish counting using image processing. International Conference on Computational Intelligence and Software Engineering, pp. 1–5.

Watson, D., Harvey, E., Anderson, M., Kendrick, G., 2005. A comparison of temperate reef fish assemblages recorded by three underwater stereo-video techniques. Mar. Biol. 148 (2), 415–425.

Yu, J., Wong, W.-K., Hutchinson, R.A., 2010. Modeling experts and novices in citizen science data for species distribution modeling. Proceedings of the 2010 IEEE International Conference on Data Mining. ICDM'10. IEEE Computer Society, Washington, DC, USA, pp. 1157–1162 (URL http://dx.doi.org/10.1109/ICDM.2010.103).