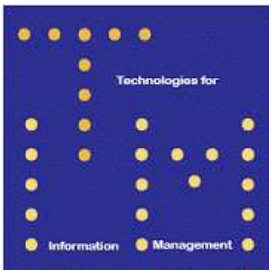
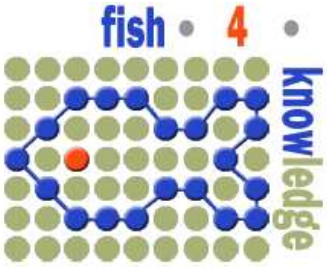




Fish4Knowledge Publishable Summary - Year 2

1 Project Context

The study of marine ecosystems is vital for understanding environmental effects, such as climate change and the effects of pollution, but is extremely difficult because of the inaccessibility of data. Undersea video data is usable but is tedious to analyse (for both raw video analysis and abstraction over massive sets of observations), and is mainly done by hand or with hand-crafted computational tools. Fish4Knowledge will allow a major increase in the ability to analyse this data: 1) Video analysis will automatically extract information about the observed marine animals which is recorded in an observation database. 2) Interfaces will be designed to allow researchers to formulate and answer higher level questions over that database.

The project will investigate: information abstraction and storage methods for reducing the massive amount of video data (from 10E+15 pixels to 10E+12 units of information), machine and human vocabularies for describing fish, flexible process architectures to process the data and scientific queries and effective specialised user query interfaces. A combination of computer vision, database storage, workflow and human computer interaction methods will be used to achieve this.

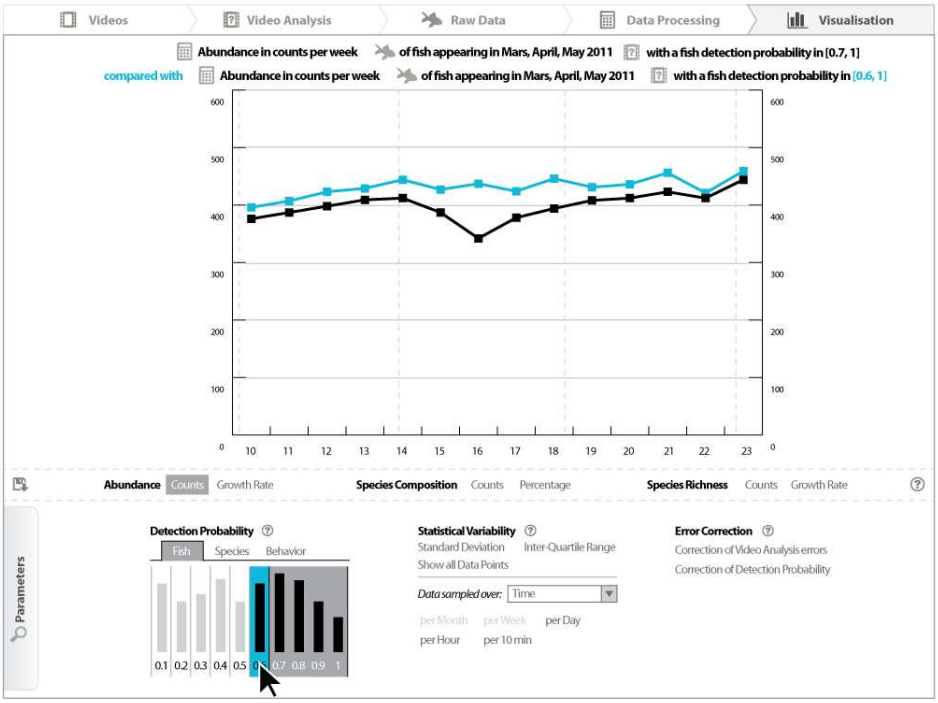
2 Publishable Project Summary

Based on the past year of scientific and engineering work, we now have the first full prototype of the data collection system:

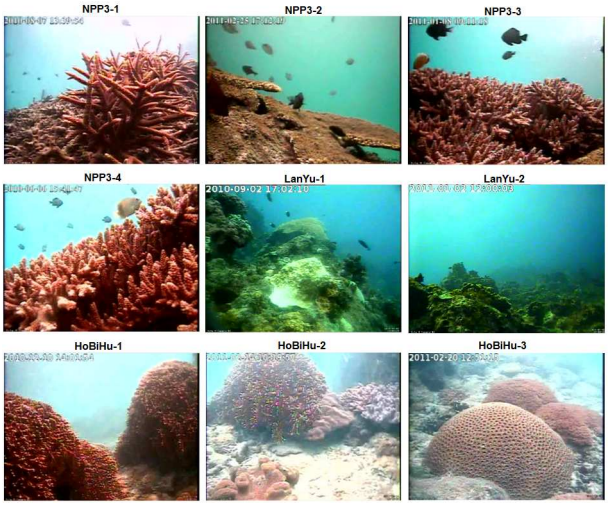
- Between 5-10 working undersea cameras (with variations due to equipment failures), leading to 91K hours of recorded video from 4 different sites.
- A working fish detection and tracking system, which has processed 2345 hours of video, resulting in 42 million fish detections which were tracked to form 4.6 million fish trajectories.
- A working fish recognition process, currently recognising the 10 most common species (representing 95+% of observed species), with an accuracy of 90% based on the current ground truth. The recognition process has been applied to 2.1 million of the tracked fish so far.
- Supporting the recognition algorithms is a ground truth dataset of 27470 verified fish images from our cameras spanning 35 species.

- The data collection and analysis components have been installed on the project supercomputer at NARL, which has been extended to 96 dedicated processors. We have also had an opportunity to use up to 1000 processors on the main NARL supercomputer (which identified a SQL database update server bottleneck).
- The main result SQL database has been relocated to NARL and is currently being extended as new videos are processed.
- The on-demand workflow system has been designed and implementation is in progress.
- The marine biologist User Interface has been designed and implementation is in progress.

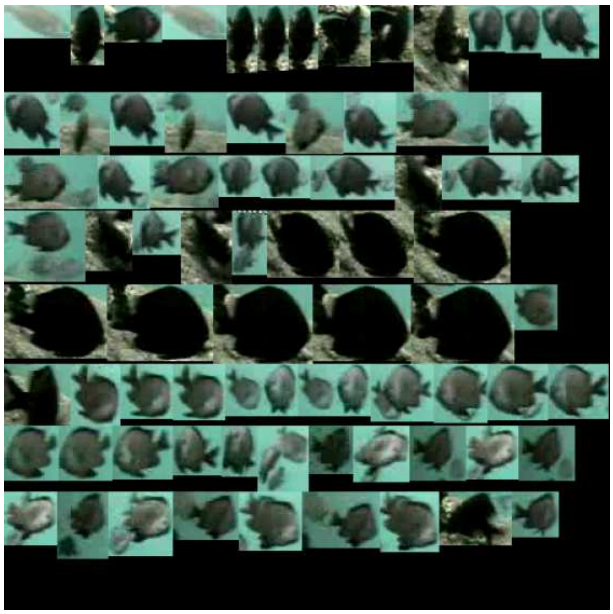
Based on visits to and discussions with several marine biologists, we have identified a set of 20+ questions to focus on, which has led to an analysis of their and the system's information needs. A mockup of the planned user interface is shown here:



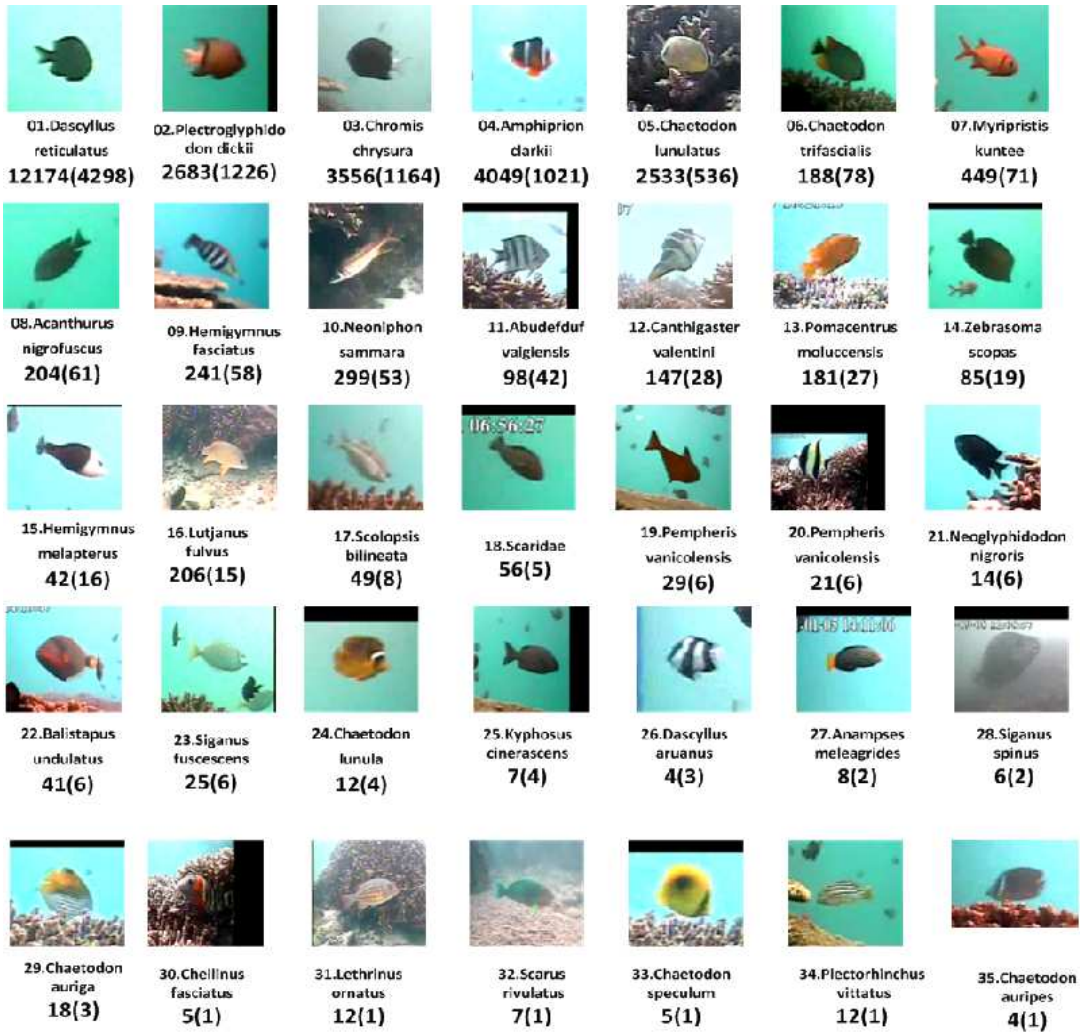
The data sources are in place: 9 cameras at 4 different locations (although only 4 are currently recording due to equipment failures and changes in local site polices). Here are views from the 9 cameras. To the right is one box of the 96 processor system which has been assembled at NARL for executing the project code.



Some example fish detections are:



The current top 35 ground truth species and the number of instances (number of trackings) of each species are shown here. 27470 detections of 8780 fish are manually labelled and verified by the marine biologists:



35 species 27470 fish (8780 trajectory)

Based on the scientific and engineering work, the project has now produced 31 publications. These are summarised at the project web site at `www.fish4knowledge.eu`, which has received 4700+ hits.

The project has organised 5 workshops this year, is leading 3 journal special issues, and is collaborating with the AQUACAM external organisation over use of the technical capabilities developed in the project.

The main objectives for year 3 will be to:

- Enhance the detection and tracking algorithms.
- Extend the species recognition algorithm to more species and higher accuracy.
- Complete system integration (workflow and user interface)

- Evaluate system performance
- Enhance system to increase data analysis and query answering speed
- Evaluate usability by marine biologists
- Catch up with all previously recorded videos.

3 Expected final results

The project will use live video feeds from up to 10 underwater cameras as a testbed for investigating more generally applicable methods for capture, storage, analysis and querying of multiple video streams. We will collate a public database from 2+ years containing video summaries of the observed fish and associated descriptors. Expert web-based interfaces will be developed for use by the marine researchers themselves, allowing unprecedented access to live and previously stored videos, or previously extracted information. The marine researcher interface will also allow easy formulation of new queries. Extensive user community evaluations will be carried out to provide information on the accuracy, ease and speed of retrieval of information.

4 Project public website

`www.Fish4Knowledge.eu`

