Fish4Knowledge Deliverable 6.6

Public Query Interface

Principal Author:	E. Beauxis-Aussalet, E. Arslanova, L. Hardman
Contributors:	CWI
Dissemination:	PU

Abstract:
This document presents the public query interface tailored to marine biologists. With this interface, users are able to browse videos and explore the video analysis results. Open-ended study of video analysis results can be performed with parameters specified by the user, such as camera location or time periods. We report our interactions with the target users, i.e., marine biologists and ecologists, with respect to the design and evaluation of the interface.

Deliverable due: 30 Month

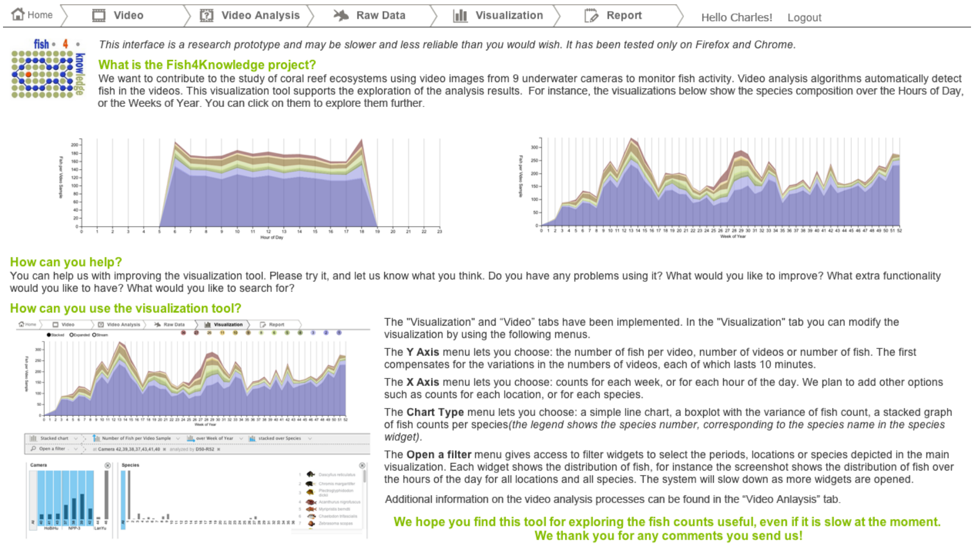

Figure 1: The Home Page of the public query interface.

1 Public query interface

The Fish4Knowledge (F4K) interface provides access to the data generated by the video analysis components. It is tailored to scientists studying fish populations in particular and, more generally, to researchers studying marine ecology (e.g., ecotoxicologists). The interface supports flexible data visualizations, and functionalities for data exploration and video browsing defined in [4, 6].

The interface consists of 5 tabs reflecting different levels of information processing within the F4K project: *Video*, *Video Analysis*, *Raw Data*, *Visualization* and *Report*. The following sections describe the usage of the *Home Page* and *Tabs* of the user interface, and can serve as a brief tutorial of the public query interface. The interface is available at these url *http://gleoncentral.nchc.org.tw*[1]. It was tested from Chrome only. Access to the F4K data will be made public in April 2014, 6 months after the F4K project ends. In the meantime, the public query interface will provide an access to a limited subset of the video analysis data.

1.1 Home Page

The Home Page of the public query interface contains i) sample visualizations of the video analysis results, and ii) explanations of the F4K tool, its video analysis results, and the means to visualize them (Fig. 1). The sample visualizations provide an overview of the *species composition* for the monitored areas. A link is provided to explore them in the *Visualization* tab. The explanations provided about the F4K tool include a section dedicated to the exploration of the uncertainties of video analysis results (e.g., missing video samples, video quality, video analysis errors). They guide users in understanding the uncertainty factors to take into account for assessing scientific conclusions from the video analysis results.

[1]This public query interface is localized on Taiwanese servers, providing a faster access from Taiwan, where our primary users are. For a faster access from Europe, a copy of the system is available at this url: http://f4k.project.cwi.nl, but it may not provide up-to-date video analysis data.

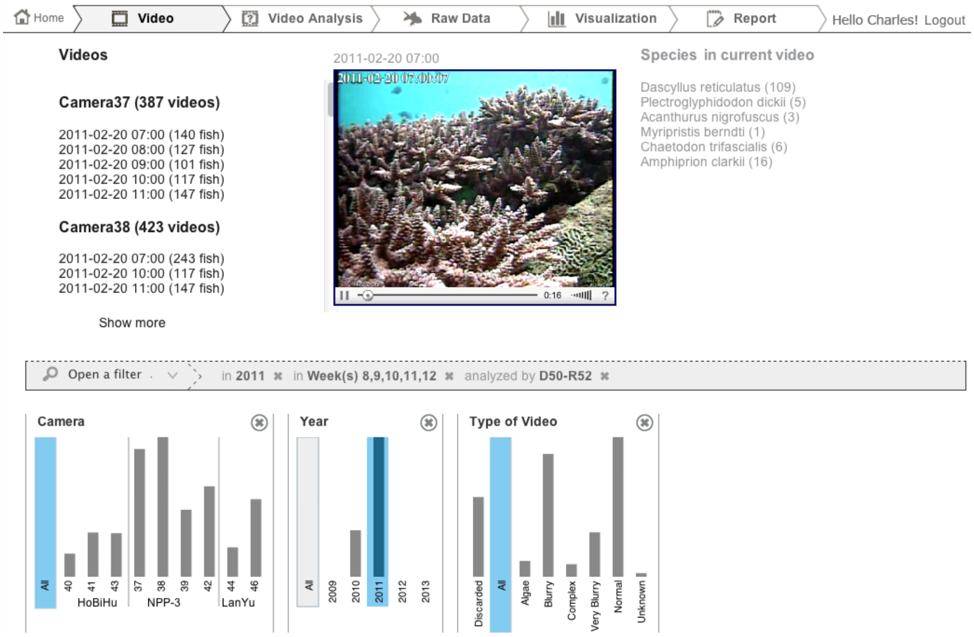

Figure 2: **The Video Tab**. In this example, the user is browsing videos from Weeks 8-12 of 2011, from all cameras, and all video quality, and analyzed by the software version *D50-R52*. These settings are shown by the *filter widgets* displayed in the lower part of the interface, and by the *filter summary* displayed above the widgets. The *filter summary* indicates only the filters that are not set to ”All”.

1.2 Video Tab

The Video Tab supports the browsing of video samples (Fig. 2). It is organized in 2 parts. The upper part contains a video player, a list of selected videos, and indications of the fish appearing in the videos. The videos in the list are selected using filtering functionalities contained in the lower part of the sub-tab. Users are provided with *filter widgets* to select:

- *Year, Week of the Year, Hour of the Day*: the time period when videos were recorded[2].
- *Camera*: the cameras from which videos were recorded.
- *Video Quality*: videos with specific image quality (e.g., Algae, Blurred, Very Blurred, Complex Background, Normal).
- *Software*: this filter lets users select videos that were processed by specific versions of the video analysis components.
- *Certainty Score*: this filter let users select videos that contain fish with specific certainty scores (which represent how a fish appearance is similar to a fish model built from training video analysis components on ground-truth fish images). To reduce user cognitive load, and since the concept of certainty score is complex for non-experts ([3]), we provide the certainty scores for Species Recognition only. Certainty scores for Fish Detection and Tracking are not displayed.

[2]More filters may be useful (e.g., calendar, lunar month), but were not implemented because of time constraints.

The video player displays the 10-minute videos collected during the project. It cannot show excerpts focusing on specific fish (e.g., from specific species, or with specific *certainty scores*). This functionality is of interest, but technical constraints prevented its implementation within the resources of the project.

1.3 Video Analysis Tab

The *Video Analysis* tab provides information about each video processing step, and evaluations of the video analysis accuracy. It contains 4 sub-tabs:

- The **Overview** sub-tab (Fig. 3) provides explanations of the 4 main video processing steps.
- The **Fish Detection** sub-tab (Fig. 4) provides the performance evaluation of the Fish Detection component. The ground-truth based evaluation is visualized in a simplified manner, so as to be easier to understand for users with no image processing expertise (e.g., no mention of True Negatives, no rates such as False Alarm Rate, nor threshold-based evaluations)[3]. Further performance evaluations are given for each of the video qualities: Algae, Blurred, Very Blurred, Complex Background, Normal or Unknown. This allows the comparison of potential errors, and the estimation of potential biases due to varying video quality.
- The **Species Recognition** sub-tab (Fig. 5) provides the evaluation of the Species Recognition component. The ground-truth based evaluations are provided with simplified visualizations, as for the Fish Detection evaluation. Further evaluations are given for each species. This allows the comparison of the potential errors involved for each species, and the estimation of potential biases amongst species. Evaluations are given for 2 versions of the Species Recognition components: one can recognize 15 species, and another another can recognize 23 species. Recognizing more species may involve decreasing the accuracy of some species. With this sub-tab, users can decide which component to use depending on the species of interest.
- The **Workflow** sub-tab (Fig. 6, [7]) allows on-demand video processing. Users can request the analysis of specific videos (from user-defined time periods and cameras) with specific component versions (e.g., with the best accuracy for the species of interest). For instance, it may consist of processing videos needed for the current visualization, but that were not yet analyzed by a specific F4K component. Users can choose the version of the components that offers the best accuracy for their particular study (e.g., a version of Species Recognition that is the most accurate for specific species). By default, the preselected components, periods and cameras are those of the current dataset displayed in the *Visualization* tab. Users can modify these parameters, and ask the workflow to estimate the time needed to complete the desired video processing. When a query for video processing is launched, users can follow its execution on the right part of the screen, and potentially cancel the query if no longer relevant.

[3]Our study [3] highlights that this level of detail is already difficult to understand, and that providing full ROC evaluation, as commonly done in the image processing domain, is likely to overwhelm users. However, omitting them completely would mislead users in their understanding of the video analysis results.

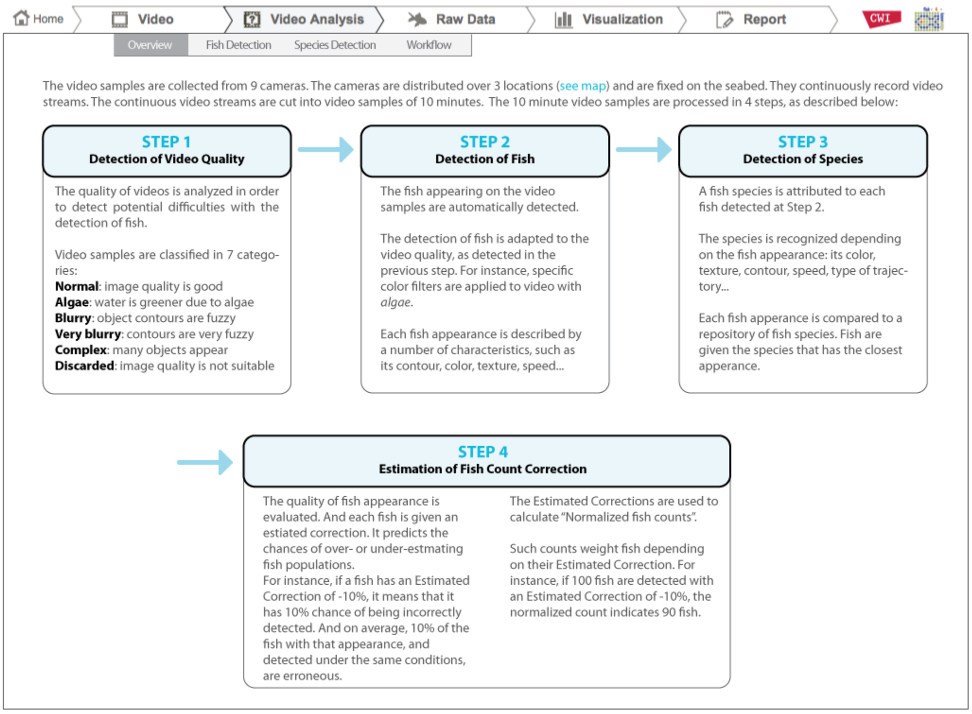

Figure 3: **The Video Analysis Tab - Overview sub-tab** provides explanations of the video processing steps.

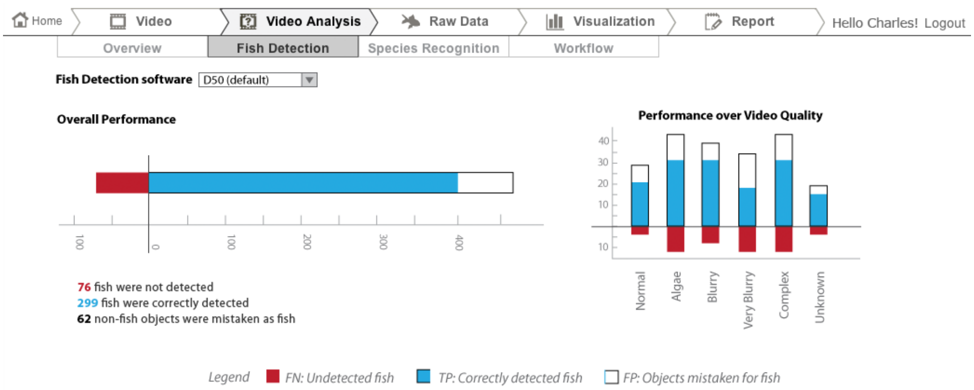

Figure 4: **The Video Analysis Tab - Fish Detection sub-tab** provides simplified visualizations of ground-truth based evaluation of the Fish Detection component. Evaluations are provided for each video quality.

4

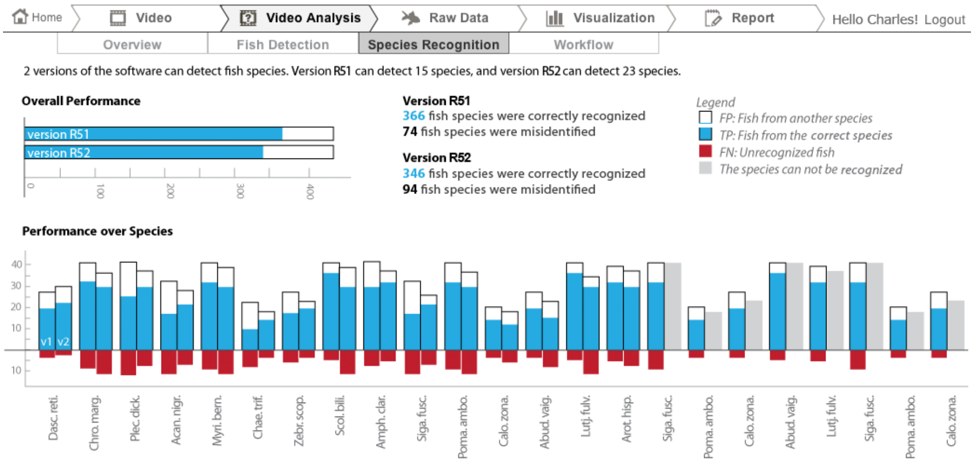

Figure 5: **The Video Analysis Tab - Species Recognition sub-tab** provides simplified visualizations of ground-truth based evaluation of the Species Recognition components. Evaluations are provided for each species.

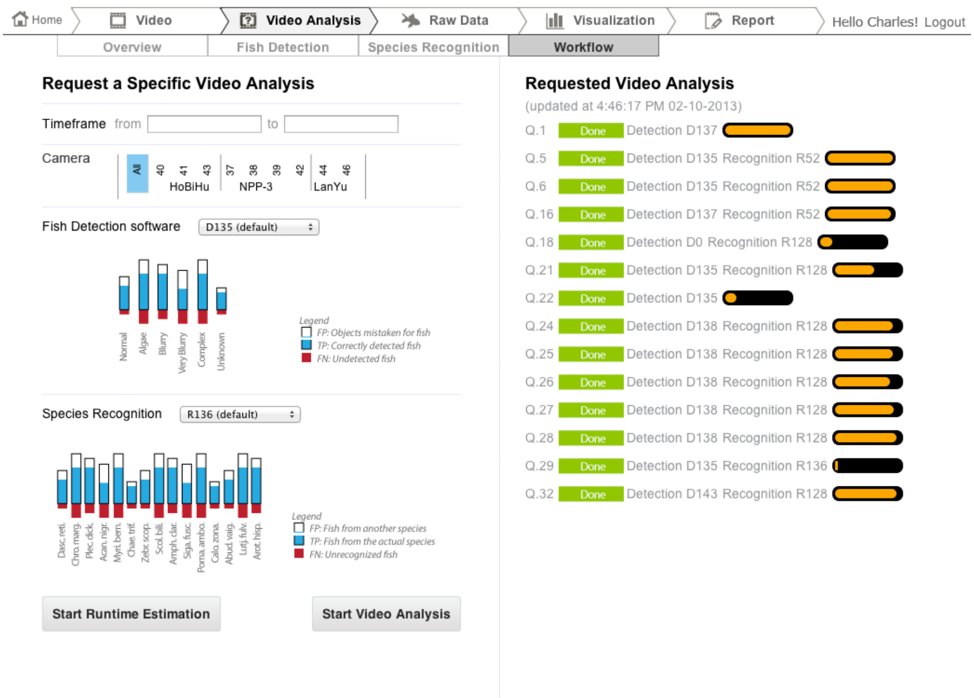

Figure 6: **The Video Analysis Tab - Workflow sub-tab** supports user requests for specific video analyses. The ground-truth based evaluation of the component versions users plan to use are shown.

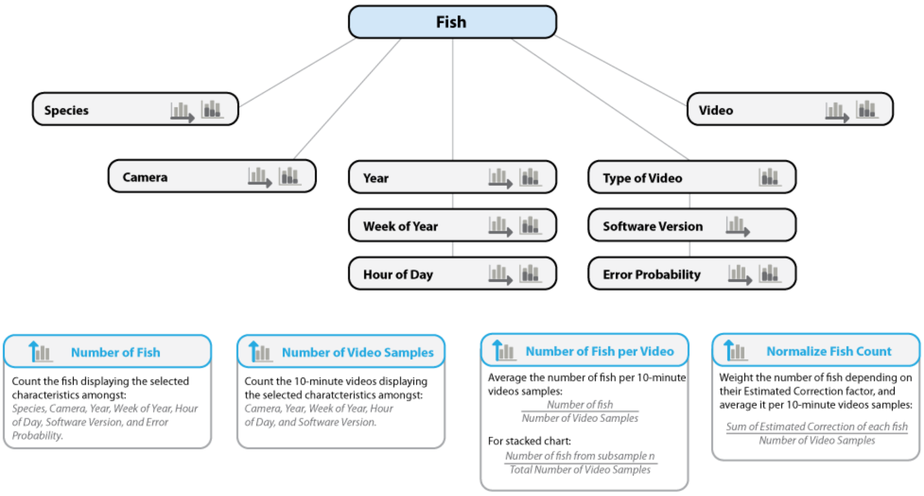

Figure 7: **The Raw Data tab** provides a schema of the video analysis data, and explanations of the Y axis metrics.

1.4 Raw Data Tab

The *Raw Data* tab (Fig. 7) provides an overview of the available video data and their properties. It helps users in identifying the information to select for their particular study. It shows all the characteristics of fish populations. They can be used to filters data. Visualizing these fish characteristics support users in understanding what fish populations they can study. It also explains the computations performed when calculating the metrics displayed in the y-axis (e.g., Number of Fish, Number of Video Samples, Number of Fish per Video Sample, Normalized Fish Count).

1.5 Visualization Tab

The *Visualization* tab supports user-defined flexible visualizations of video data, including visualizations that address trust issues (e.g. counts of processed videos or flawed video quality). For instance, Fig. 8 shows a visualization of the variations of fish counts over the hours of the day. **Zone A** of the interface contains the main visualization that displays the fish counts. **Zone C** contains *filter widgets* that supports:

- The selection of the dataset of interest (e.g., the timeframe, location and species of interest).
- The overview of fish counts that can be obtained with different parameters (e.g., fish counts per location, per species, per year, etc).

More filter widgets can be opened on-demand (e.g., to select species of interest, or data from a specific versions of the video analysis software). All the filter widgets provided by the public query interface are given in Fig. 9.

Zone B of the interface allows users to specify what is represented by the axes of the main visualization. For instance, while the y-axis represents fish counts, the x-axis can represent the distribution of such counts over weeks of the year, hours of the day, or locations (i.e., cameras).

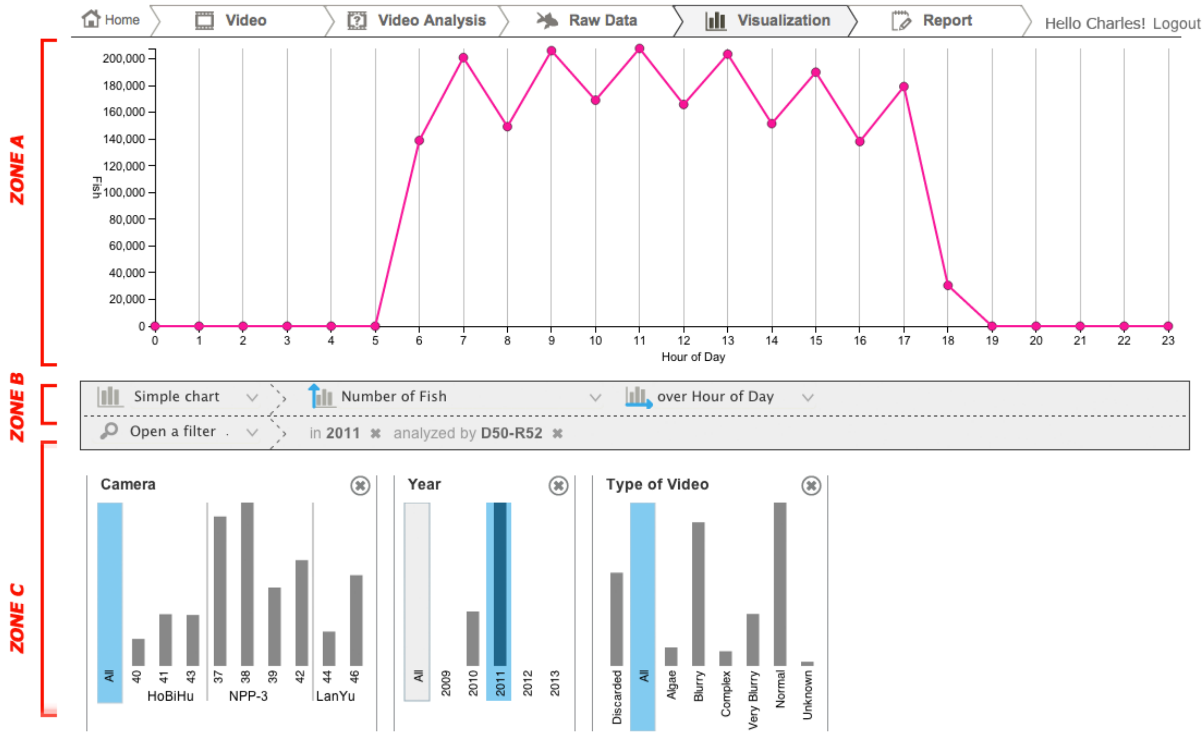

Figure 8: The **Visualization tab** displaying a visualization of Number of Fish distributed over the Hours of the Day. The *Zone C* displays the distribution of fish counts over Cameras, Years, and Video Quality. The visualizations concern the dataset produced by *component D50-R52* and for the videos collected during the *year 2011*.

Users can select the type of graph (e.g., stacked chart, or box plot), which leads to the display of dedicated menus for adapting the visualization further. For instance fish counts can be stacked by species, camera (Fig. 10), or time periods such as year (Fig. 14). The subsampling of fish counts per dimension, such as species, camera or year, can be specified using a drop-down menu. The menus that allows to modify the main graph contain the following options:

- **Y axes** of the main graph, and of the *filter widgets*: Number of Fish, Number of Video Samples, Number of Fish per Video Sample, Normalized Fish Count[5], Number of Species.
- **X axes** of the main graph: Year, Week of Year, Hour of Day, Camera, Species, Certainty Score, Estimated Correction[5], Video Quality, Component Version.
- Subsampling for **stacked charts**: Year, Hour of Day, Species, Camera, Certainty Score, Estimated Correction[5], Video Quality[6].
- Subsampling for **boxplots**: Year, Week of Year, Hour of Day, Species, Camera, Certainty Score, Estimated Correction[5], Video Quality, Video[6].

This direct interaction with the axes of the graph was easy to understand for users, while offering a large choice of visualizations and flexible data analyses. It satisfies a wide range of user needs, in a context where biologists need very different data analyses depending on their specific research goals.

In addition to querying for fish count data, users can also query for data related to the number of video samples that are available, i.e., the number of 10-minute videos from which fish observations are automatically extracted. Although cameras continuously record videos as long as there is daylight, the number of video samples may not be the same for all cameras and periods of time. Video samples may be unprocessed, being in the workflow queue for their processing to be executed later, or may be discarded due to encoding errors. Fig. 11 shows the number of video samples from which the fish counts in Fig. 8 and 10 were extracted.

The variations in the numbers of video samples have a direct impact on the analysis of fish counts, specifically, the more video samples, the more fish. Our tool provides means to compensate for these variations. Users can visualize the average number of fish per video sample, as shown in Fig. 12. However potential biases still remain, since the risk of under- or over-representing fish is higher when the number of samples is smaller. One way of compensating for this is to ensure that the number of samples is the same, so that potential under- or over-representations remain the same over the whole time period and locations of study. This condition would then allow biologists to draw conclusions on the trends that can be observed in the fish counts. For this reason, our system supports user queries for on-demand video processing, i.e., in the *Workflow* sub-tab. This functionality is useful for cases where i) the number of video samples is uneven; ii) the number of video samples is too small.

In addition to the uncertainty due to varying numbers of video samples, others factor of uncertainty concern the misidentification of species, and the impact of video quality (e.g., *Blurred* or *Normal* videos) on fish detection and species recognition. Some species are more difficult to recognize, and some video quality may produce more errors, leading to over- and

[5] This metric is experimental and is currently under implementation. It may not available by the end of the project, but its study will be continued afterwards.

[6] Some options may be blocked depending on the over options chosen for the Y and X axes. For instance, the Numbers of Species cannot be stacked per species.

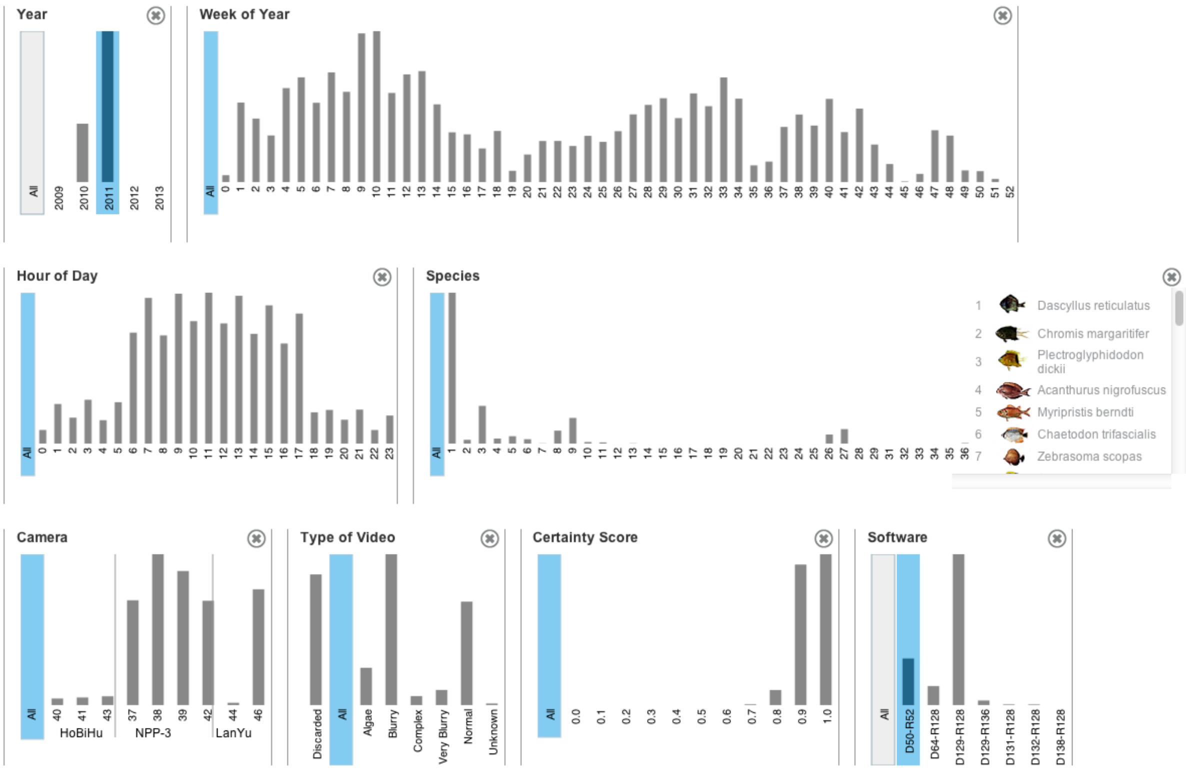

Figure 9: The *filter widgets* that allow the selection of the dataset of interest, i.e., the time period, location and other characteristics of the fish to study, as well as the versions of the software that produced the data. The histograms of the widgets provide an overview of various data distribution. The Y axis of the histograms represents the same metric as for the main visualization. This example shows the distribution of fish counts over several dimensions. We can see that only 4 versions of the software can provide fish counts for the periods and locations selected by users.

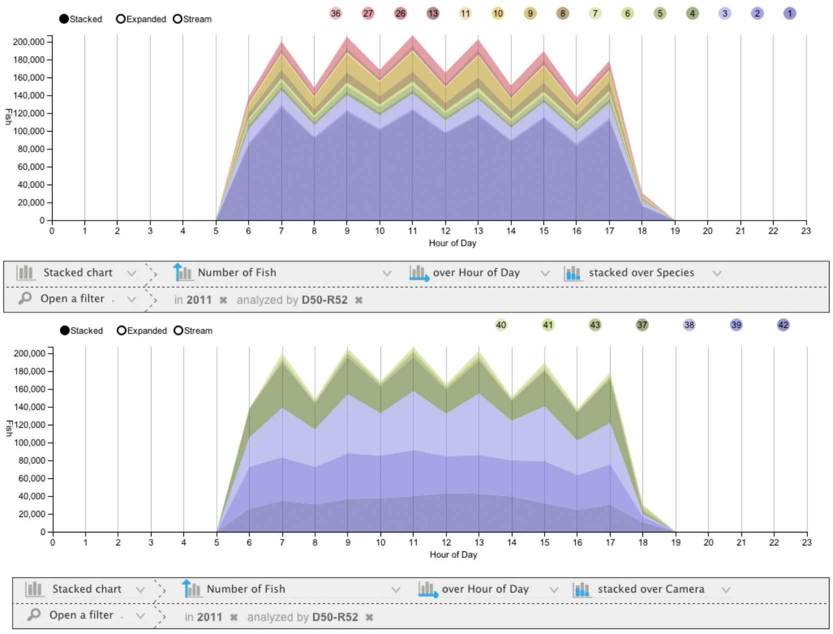

Figure 10: Visualizations of the same fish counts (as in Fig. 8) stacked by species, or by camera.

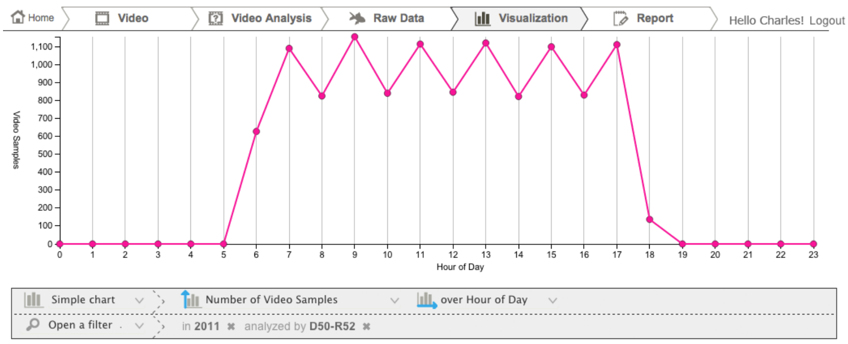

Figure 11: A visualization of the number of video samples that are processed, and from which the fish counts in Fig. 8,10 were extracted.

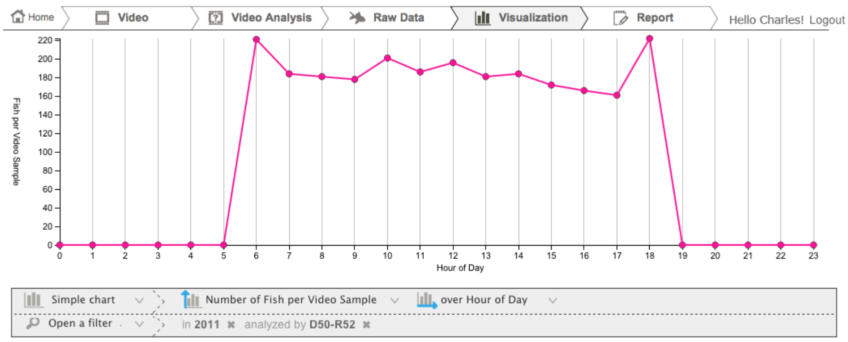

Figure 12: A visualization of the average number of fish per video sample. This compensates the variations of numbers of samples, and their impact on fish counts (e.g., variations in Fig. 8,10).

under-estimation of some fish populations. We initiated the study of an experimental metric, called *estimated correction*, that aims at compensating for such biases. This metrics takes into account the conditions in which a fish is identified (e.g., which species, video quality and certainty scores). The *logistic regression* technique[7] is used to calculate the potential under- or over-estimates of fish populations identified under these conditions. This allows the normalization of fish counts with a simple formula, e.g., a population of 100 fish with an *estimated correction* of -10% represents a normalized fish count of 90 fish. This is based on evaluations showing that with the given conditions (e.g., specific species, video quality and certainty scores) it is likely that fish are over-estimated by 10%. Similarly, a fish with an *estimated correction* of +20% means that with the given conditions it is likely that 20% of the fish were not detected, giving an estimated count of 1.2 fish per fish. Such normalization potentially correct the biases due to over- and under-estimation of some species. Normalizing fish count potentially reduces the video analysis uncertainty. However, normalized fish counts do not suppress the uncertainty, nor represent the actual number of fish with full certainty.

The computation of *estimated correction* is an experimental functionality currently under-development and is not likely to be available by the end of the project. For instance, more ground-truth is needed for each video quality (e.g., Blurred, Algae, Normal), and more time is needed for testing and evaluation. Until this has been implemented, we provide *certainty scores* for estimating the video analysis uncertainty. In any case, we will continue with investigating fish count normalization after the end of the project.

1.6 Report Tab

The *Report* tab (Fig. 13) supports manual grouping and annotation of visualizations created in the Visualization tab. Visualizations can be added to and removed from a *report*, and their interpretation can be described with a visualization title and a comment. Using the *Download* button, users can save the report they are currently working on. Downloaded reports consist of a text file containing a list of parameters. They can be stored or sent to other biologists as any kind of text file. To visualize a downloaded report, users can upload the parameter files with the *Upload* button of the *Report* tab.

2 User scenarios executed with the public query interface

This section illustrates how the public query interface addresses user information needs, as successively investigated in D2.1 [8], D2.2 [5] and D2.3 [6]. The user scenarios developed in D2.2 [5] serve to illustrate which functionality of the public query interface addresses which information need.

The *Charles* scenario describes tasks related to the study of fish population dynamics, e.g., the variations of fish abundance over time or location. The 8 steps of the scenario are addressed as follows, where at each step of the scenario, we indicate which user requirements are addressed, using the requirements summarized in D2.3 - Appendix I [9].

1. **Yearly counts of fish**: *Charles* has several options to visualize fish counts per year, as shown in Fig. 14. Such visualization addresses the requirement **D2.1-A** "Support the

[7] http://en.wikipedia.org/wiki/Logistic_regression

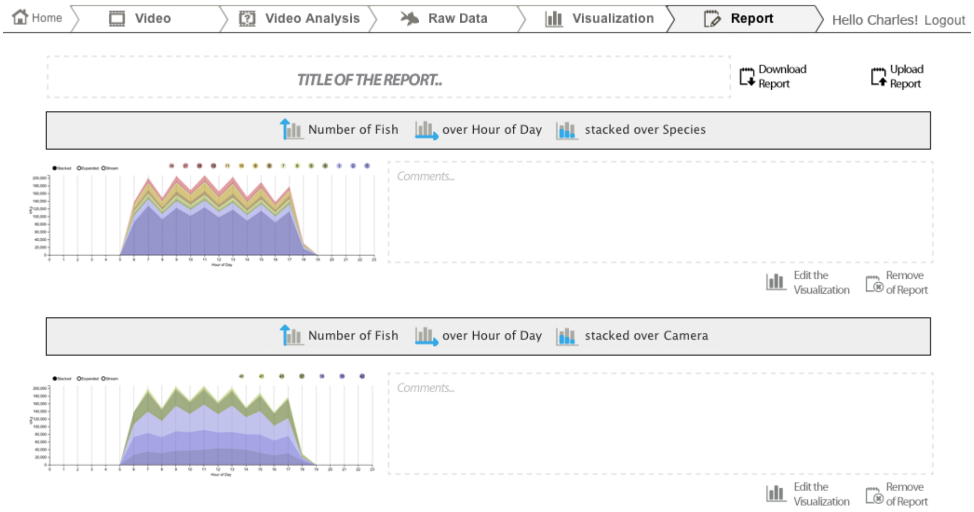

Figure 13: The **Report tab**.

analysis of population dynamics".

2. **Counts of fish for a specific month**: *Charles* can zoom in yearly fish counts, and visualize counts for a specific month, as shown in Fig. 15. This addresses requirement **D2.1-A**.

3. **Control the current visualization**: from the Visualization tab, *Charles* can control which video samples produced the fish counts, as shown in Fig. 16. Fig. 9 shows the widgets used for controlling which software component produced the video data, and which subset of the data is visualized. Finally, *Charles* has access to more detailed information through the other tabs. These visualizations address the requirement **D2.3-A** "Expose the uncertainty of video analysis components".

4. **Control the videos**: to complement the controls provided by the *Visualization* tab, *Charles* can browse the videos in the Video tab (Fig. 2). It addresses the requirement **D2.1-B** "Support the browsing of videos of interest".

5. **Control the video analysis components**: in the *Video Analysis* tab, *Charles* can control the versions of the Fish Detection and Species Recognition components, and the potential errors they imply for specific species or video quality (Fig. 4-5). This guides the selection of the component versions to use for studying fish populations. This addresses the requirement **D2.3-A**.

6. **Control the selected dataset**: *Charles* can overview the dataset characteristics in the *Raw Data* tab (Fig. 7). It shows that a threshold can be applied on *certainty score*, and eventually on the *estimated correction* need for normalizing fish counts (e.g., to discard fish population needing an important correction). For instance he can filter out fish with high chances of errors. This tab also shows which metrics can evaluate fish abundance while taking into account specific uncertainty factors. The *Number of fish counts per Video Samples* compensate for varying numbers of video samples. And the *Normalized fish count* compensate for both the varying numbers of video samples, and the species

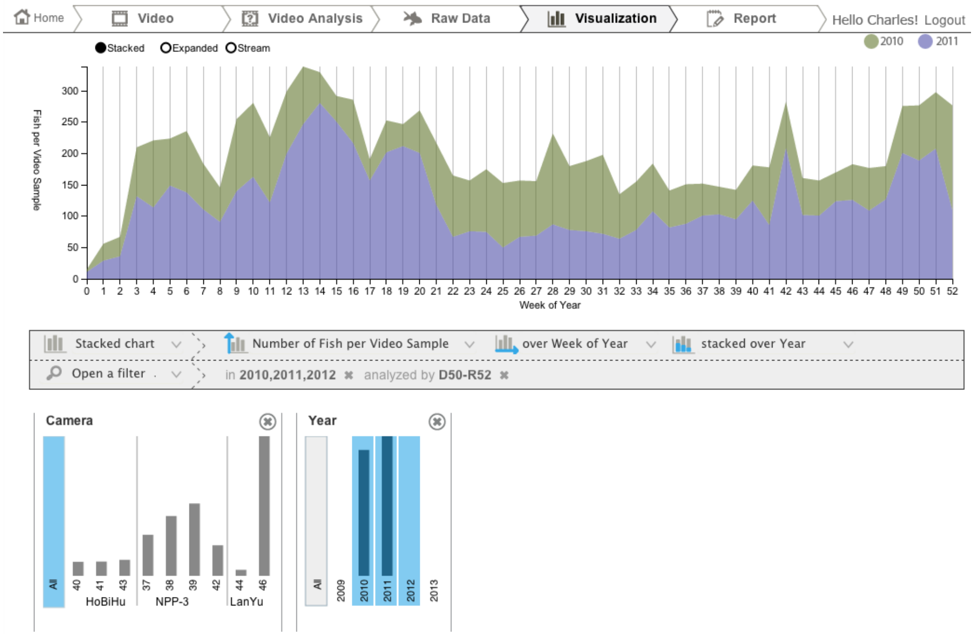

Figure 14: A Visualization of annual fish counts.

misidentification. This addresses the requirement **D2.3-B** ”Estimate the errors contained in the visualized datasets”.

7. **Fish counts per location**: *Charles* can visualize fish counts per location, as shown in Fig. 10. These visualizations address the requirement **D2.1-A** ”Support the analysis of population dynamics”.

8. **Comparing fish counts per area**: *Charles* can compare localized fish count by using stacked charts, as shown in Fig. 13, and/or by gathering visualizations for specific areas using the *Report* tab. This addresses the requirement **D2.1-C** ”Support the identification of trends and correlations of trends”.

The *Erica Scenario* (studying the impact of environmental conditions such as typhoon) was not completely implemented due to technical constraints. However, environmental events can still be studied using the public query interface. Users can manually select the periods and locations where environmental events occurred, and analyze their impact on fish abundance. Further, the *Report* tab supports comparisons of fish populations before, during or after environmental events.

3 Promoting the results to the marine biology community

3.1 User Studies with Marine Biologists

To develop the F4K interface we conducted a series of 2 user studies involving a total of 34 biologists from Taiwanese and Dutch institutions. We also conducted 3 system demonstrations in Taiwan. The studies were conducted at different stages of the design process and, therefore, used different prototypes. We used interviews, questionnaires and think-aloud methods to elicit

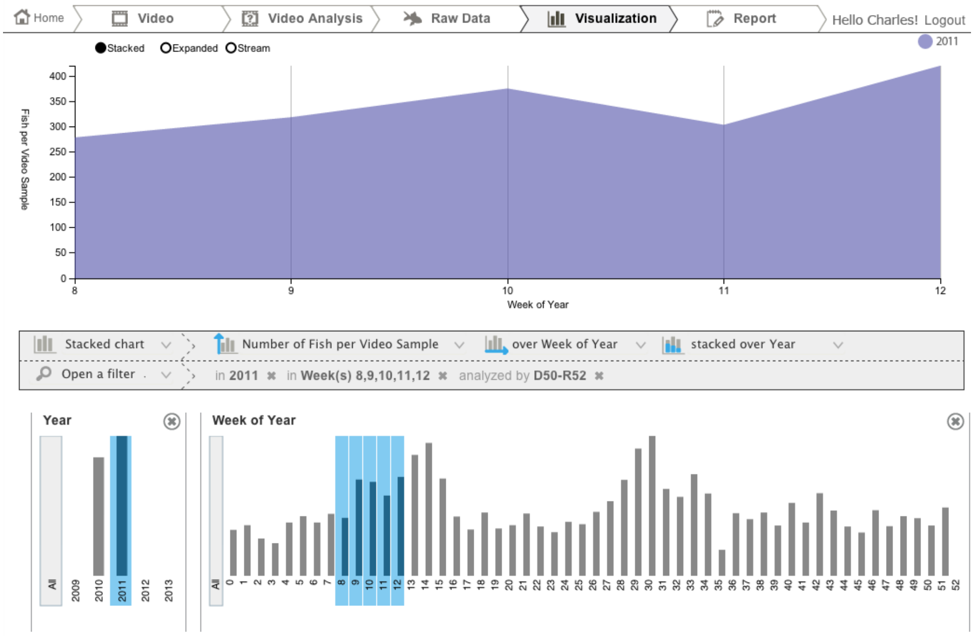

Figure 15: A Visualization of fish counts from March 2011.

Figure 16: A Visualization of the numbers of video samples available for March 2011. The lower part of the interface shows the distribution of videos per Camera, Year, and Type of Video Quality.

both explicit and implicit expert feedback on the software. Our semi-structured interviews addressed questions such as the acceptance of the video analysis tools in general and F4K in particular, the methods to validate the results produced by F4K computer vision components, and the advantages and potential improvements of our tool.

Trust and Data Provenance Study

This study addressed the support of user trust in the video analysis result by providing provenance information. We explored the level of detail provided about ground-truth evaluation. We successively introduced marine biologists to ground-truth (i.e., man-made) fish counts, to True Positives, False Negatives and False Positives, and finally to certainty score thresholds. Each time new technical concepts were introduced, we measured i) user understanding of the concepts, ii) user trust in the video analysis results, iii) user acceptance of the tool for scientific research, and iv) the satisfaction of user needs for provenance information. This study is reported in [3]. Our results show that the technical concepts of ground-truth based evaluations are complex to understand for non-experts. Participants were overwhelmed, even if not exposed to further concepts such as True Negatives, and rates such as those used for ROC curves.

Situation Awareness Study

While working with our complex data analysis system, it is essential for users to maintain a good understanding of the available data, and of the aspects of the data that influence the validity of their scientific study. Moreover, the complexity of tasks could contribute to tunneling user attention and lead to human mistakes (e.g. because users are focused on their task they may not notice that they visualize the wrong y-axis, or use unnecessary filters). We evaluated how the F4K interface supports the performance of users depending on task complexity. We asked marine biologists to carry out standard data analysis tasks that could be done in the public query interface. The tasks varied in their complexity and allowed us to see what issues appear at different levels of task complexity. We observed that uncertainty issues are likely to be overlooked. For instance, most participants did not spontaneously controlled the uncertainties due to scarce video samples.

System demonstrations

We organized 3 demonstrations of the F4K public interface to marine biologists with research interests in coral reefs. We presented the functionalities of our tool and briefly described the underlying video analysis processes. After the presentation we discussed the interface with the biologists and allowed them to freely explore the F4K tool by either using it individually or asking the presenter to perform certain queries. We collected insights about the most relevant functionality of the interface, suggested improvements, and potential usage for scientific research.

3.1.1 Participants

The choice of participants was determined by the objectives of the user studies. This principle allowed us to involve biologists that study other organisms than coral reef fish (e.g. deep sea fish or plankton). We also investigated how the F4K tool could be applied in their field of research.

Coral reef fish - We interacted with three teams studying *population dynamics* and *fish systematics* of coral reef fish: Academia Sinica (Taiwan), National Museum of Marine Science and Technology (Taiwan) and Wageningen University (The Netherlands). They traditionally use diving observations to collect data. Some teams already use video cameras to collect data. For example, one team uses baited stereoscopic cameras to obtain samples from different locations. Another team dives with hand-held cameras and moves along the specific path. The teams manually analyze the videos and would potentially be interested in applying video analysis methods to avoid manual counting. To the interest of this group of researchers, the F4K tool can provide information about detected species, and potentially about fish behavior.

Open sea fish - We interacted with a team of researchers from a Dutch institution that study population dynamics (e.g., abundance and distribution) of pelagic fish, i.e., fish living in the open sea. This team traditionally uses commercial and experimental fishing to collect data.

Deep sea fish - We interacted with a team of researchers from a Dutch institution that study deep sea ecosystems. They study *abundance*, *distribution* and *biodiversity* of benthic organisms (i.e., living on the sea bed) of the North Sea deep trenches. They record videos by attaching a lighted cameras to a vessel, and hovering the sea bed. They calibrate the distance between the camera and the sea bed, in order to measure the body size of organisms.

Corals - We interacted with several teams based in the National Museum of Marine Biology and Aquarium (NMMBA) in Taiwan, and studying various aspects of coral ecology: *reproduction*, *physiology*, *restoration* of corals, and *biodiversity* of coral reef ecosystems. They have well accepted and widely used services for semi-automatic identification and monitoring the growth and health rate of corals. They traditionally use static pictures obtained while scuba diving to calculate coverage and identify corals morphology. To collect coral reef data from different parts of the world, they crowdsource data collection by uploading underwater pictures to an online service for automatic recognition of coral species [1]. The combination of F4K capabilities and currently used coral reef monitoring services could provide biologists with a powerful tool to monitor the whole coral reef ecosystem.

Anatomy, physiology - We interacted with two teams based in NMMBA and in Academia Sinica whose research is focused on *ecotoxicology*, *animal behavior, anatomy, physiology*, and *sense organs*. Ecotoxicology is the study of the effects of manufactured chemicals and other anthropogenic and natural materials and activities on aquatic organisms at various levels of organization [2]. These biologists suggested that the F4K tool *"would be very useful in ecotoxicology studies on monitoring fish population and community at sites with different contamination levels"*.

Plankton, microorganisms - We interacted with researchers from NMMBA studying plankton and microorganisms. These organisms are the base of all food chains and a key component of coral reef ecosystems [10]. We investigated whether the F4K tool could provide valuable information for their research. The F4K video analysis technique could be used to identify plankton preserved alive within sea water samples. This overcomes the issue of analyzing dry plankton, which lost their original fragile shape.

Marine biology education - The team of biologists based in NMBBA monitor fish population and species composition in the aquarium to maintain the healthy balance of the artificial marine ecosystem. This aquarium is also a platform where the marine researchers conduct studies. The F4K tool could support above activities by providing automatic count of fish, species and detecting their behaviour over a long period of time. The team also creates educative interactive programs to bring awareness to the Marine Aquarium visitors. The NMMBA could potentially benefit from the F4K tool redesigned as an interactive game to learn fish species and monitor their activities.

3.1.2 User Feedback

Video analysis tools are relatively recent in this community and no well-accepted data analysis framework has been set up for the usage of video data for marine biology research. Our user studies provided valuable insights for understanding the potential usage of our tool, and, more generally, for understanding the acceptance of video analysis tools by the marine biology community.

The types of evaluation that are well-accepted by the image processing community are not easy to understand by marine biologists (e.g., ROC evaluation). In our studies the majority of biologists encountered difficulties with understanding their technical concepts. Thus it is difficult for them to evaluate the potential errors introduced by computer vision components. We observed that users tend to overlook the technical details that can bias their analysis. They also do not perceive the software as fully reliable, and expect large numbers of errors, as well as biases (e.g., systematically larger error for specific species or video quality). However, biologists are still likely to accept the tool for their research. We identified 2 factors that support such good acceptance of video analysis tools for marine biology research. First, video analysis tools can considerably reduce the effort currently involved for manual annotation of videos. Second, biologists are used to deal with the high level of uncertainty in the collected data (e.g., fishery data, diving observations), since underwater ecosystems are difficult to access, and are often impossible to observe directly (e.g., open sea, deep sea). The most important user feedback concerns the following issues:

Provide understandable validation of the video analysis software - The technical methods used to validate the tool could be difficult to understand and accept by the marine biology community. Therefore, they suggested using methods adopted from biology (e.g., counting fish in a controlled environments, repeating measurements). They also were eager to trust the image processing expert opinion while choosing the settings for the software (e.g. the most reliable version of the software to detect particular species). Addressing this feedback led to the visualizations shown in Fig. 3-5.

Provide visualization of potential biases - Two experts requested to repeat the measurements performed during ground-truth based evaluations. Repeating measurements is a common practice in marine biology for dealing with uncertainty. The experts requested to visualize the standard deviation of True Positives, False Positives, and False Negatives, as measured during cross-validations. The experts wanted to evaluate if the levels of errors are systematically different under specific conditions (e.g., specific species and video quality). This would allow the identification of potential biases in the video analysis results. We plan to address this requirement in future work, for instance by evaluating the usability of the visualization shown in Fig. 17.

Provide comprehensive provenance information - Regarding uncertainty issues, biologists expressed requirements for technical information other than ROC-like evaluation:

- The image quality of the video samples used (e.g., fuzziness, murkiness). Video quality may bias the video analysis results. For instance, seasonal events like typhoon can influence video quality, and thus the seasonal abundance patterns observed. We addressed this issue by providing a filter widget allowing user to select datasets extracting from specific videos quality.
- The performance of the video analysis components for various video quality (e.g., more errors may occur with murky videos). We addressed this requirement by providing the visualizations shown in Fig. 4.
- The rate of duplicates of single fish in fish counts. Some species may produce more duplicates than other species, because of their natural swimming patterns. This is a potential bias for studying the relative abundance of each species (e.g., species composition). We were not able to address this complex issue, which requires the collection of more experimental data.
- Description of the habitats observed within the cameras' field of view (e.g., the species of coral). We were not able to address this requirement because we focused our resources on more important refinements. We assume that users can rely on the video browser (the *Video* tab) to investigate the habitats observed by the cameras.

Locations of the cameras - The coverage of the ecosystem of study is essential and specific to every research topic. Many biologists want to choose the location for their cameras individually. Optionally, they would like to have a service that could process videos captured by cameras independent from F4K. Such videos could be recorded in transects, e.g., with a moving background. Several biologists are interested in taking this further internationally.

High-level information needs - A number of additional visualizations and UI features were suggested by users, such as: the integration of a calendar and of lunar month for filtering datasets of interest, the usage of the traditional data analyses widely used for biodiversity research, access to detailed description of fish species (e.g., image of the species, link to fishbase.org).

3.2 The 9th Indo-Pacific Fish Conference, Okinawa, June 2013

The conference is held every 4 years and forms a focal point for researchers in the region to discuss current trends in their work. The conference has 5 parallel sessions, with a total of around 450 presentations of 15 minutes each. Of the parallel sessions, at most one was in any way relevant to the Fish4Knowledge work, and of these about one quarter to one third of the presentations were related to video analysis or long term monitoring of fish abundance, giving a total of around 30 presentations with some relation to the project.

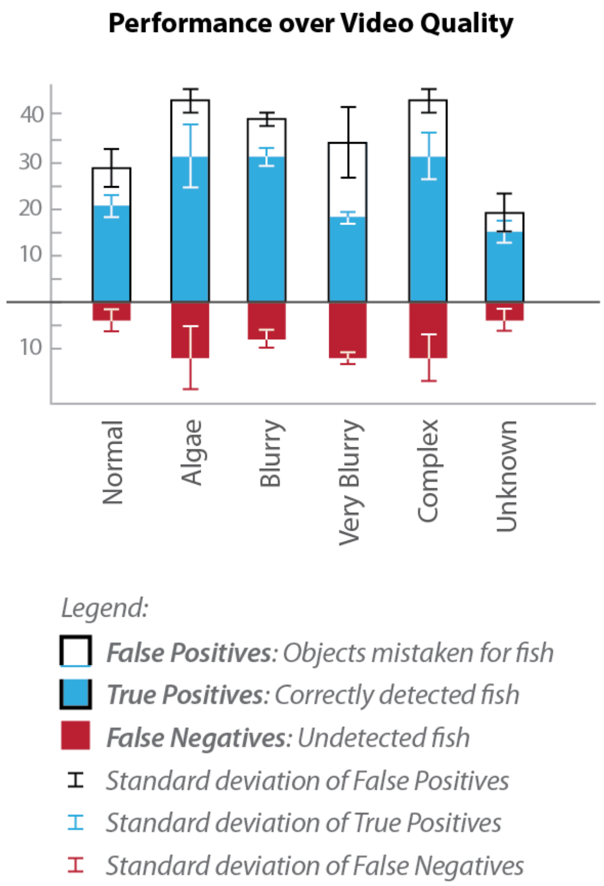

Figure 17: An experimental visualization of video analysis performance (using simulated data). It extends the visualizations of the *Video Analysis* tab by adding information about the variability of the levels of errors. It displays the standard deviation of errors measured during cross-validations (i.e., when repeating the evaluation with randomized ground-truth samples). For instance, the False Negatives (the red, lower, bars) of Algae-obstructed videos has high variability. When compared with Blurry videos, it is not likely that the levels of False Negatives are systematically different between these types of videos. On the other hand, it is likely that the levels of False Negatives are systematically different between Blurry and Very Blurry videos (the ranges of the standard deviations do not overlap). This would indicate a potential bias in the video analysis results: the results from Very Blurry videos may systematically contain more errors than the results from Blurry videos.

Lynda Hardman attended the conference on behalf of the project. She presented the project and the public query interface on the last afternoon of the conference (*Fish4Knowledge: Large scale coral reef fish monitoring using undersea computer vision methods*[8]). She approached speakers who had given presentations that had some relation to the project. She explained to them the goals and results of the project, and asked them about their own research related goals. Around 10 researchers showed keen interest in the project. She sent them further information on the project, and a link to the public version of the web interface.

The Fish4Knowledge presentation, *Fish4Knowledge: Large scale coral reef fish monitoring using undersea computer vision methods*[8], was given on the last afternoon of the conference.

A number of groups are already collecting videos for analysing fish abundance and behaviour and it is likely that many more are moving in this direction. Methods for capturing video differ. Examples are:

- Follow a number of individual fish for 3 minutes and then analyse the video for time spent on different behaviours, e.g. time spent feeding, searching, travelling, alerting (rare). (Australian National University.)
- Lower a camera into the water and leave it for a few hours (Museum of New Zealand Te Papa Tongarewa),
- Set up a camera on the sea floor (The University of Western Australia) and leave if for days, weeks or even months.

In all cases the videos are currently analysed by hand.

For abundance, the measure $MaxN$ is used, where the researcher scans through the video for the highest density of fish in a frame within a time-span of video. The fish in that single frame are counted. A commercial product is sold for around $10,000 that supports this type of (manual) counting activity. Researchers are also interested in fish body size, and use stereo cameras to obtain this information. Counting is again done manually with system support.

One researcher was worried about the accuracy of each individual detection. During the talk an explanation was given that the relative numbers of fish detected for the different certainty scores does not change the results of relative comparisons, but that the system would be unlikely to be able to give accurate absolute abundances. (Note that this is also true of the $MaxN$ measure.)

Two talks were about the effect of divers on the presence of fish (Julia Santana-Garcon, Steve Lindfield, both from the University of Western Australia). Some fish species really dislike the sound of the breathing equipment and are not seen at all by divers. This should be taken into account when doing the comparison between diver and system counts within the project. From the talks, it was clear that nothing means anything to the biologists without satisfactory significance tests. This is available using the box plot chart of the F4K tool.

Another method was used that showed clustering along different dimensions (for example, the distribution of species characteristics, such as size or age, along a water temperature gradient). The method was not explained during the talks, so that further investigation is needed to understand the statistics used to obtain the clustering.

Participants were not interested in the data we have already collected in the system: probably because it was not collected to answer their questions. They are interested, however, in how to

[8]Slides available at http://www.cwi.nl/ lynda/talks/2013/Fish4Knowledge130627.pdf

re-use or adapt the algorithms for their own videos. Questions were often about how transferable the algorithms are, e.g.,

- can the algorithm detect fish where the background is more fish, instead of the backgrounds the current algorithms were trained on;
- would the algorithm work on an all-water background.

People were interested in which species we are already able to recognise. Potential future work topics are the following:

- Create algorithms to quickly scan video sequences and identify the highest density of fish (in an example video from Julie Ann Hartup, UOG Marine Laboratory, there were hundreds of a single species in a single HD frame), and then count those individuals in that frame, giving the MaxN measure.
- Create algorithms for species detection from stereo camera images that also result in size measurements of the individual fish. For instance, the AQUACAM project, a collaboration between F4K teams and the Caribsave partnership in the Caribbean, is already pursuing this goal.
- Facilitate the creation of a community to help the technology spread by collecting algorithms for detecting fish along with corresponding sample videos and ground truth data sets.

3.3 The 48th European Marine Biology Symposium, Galway, August 2013

The symposium is held every year in a European country, and gather a variety of marine biology researchers, which represent developing areas as well as more traditional areas. The 48th symposium had themes emphasizing *biodiversity in a changing ocean*. Lynda Hardman, Emma Beauxis-Aussalet and Bas Boom attended the symposium. They presented a video[9] and a poster of the F4K project, along with a live demonstration of the public query interface. Biologists were free to use the interface on 2 demonstration computers. They collected the following feedback:

- The video analysis tool is of interest for other biologists who use cameras to collect data, but who currently manually analyze the footage.
- Video image is the best method for studying tide-swept benthos, or other seabed ecosystems that accumulate soft layers of sediments. Collecting sample pieces of organisms (e.g., with a Remote Operated Vehicle) imply moving the sediments, thus making the water turbid and negatively impacting the ecosystem. Further, these ecosystems may be too deep, or the current may be too strong for diving observations.
- A representative of the *Encyclopedia of Life* foundation[10] proposed to host the F4K data on their website, with a link to the public query interface. This would disseminate the outcome of the project, and contribute to the work of other researchers.

[9]The video is accessible on youtube (www.youtube.com/watch?v=AFV-FiKUFyI) and on the project's website (http://groups.inf.ed.ac.uk/f4k/)

[10]`eol.org`, it aims at providing *a webpage for every species*, and started with funding from the John D. and Catherine T. MacArthur and Alfred P. Sloan Foundations.

- The representative of the *Encyclopedia of Life* foundation was impressed by the public query interface and its ability to provide a wide range of data visualization, while remaining friendly and easy to use (*"Within a single page I can access everything I want to see."*).

3.4 The 9th Baltic Sea Science Congress, Klaipedia, August 2013

The congress is held every year in a Baltic country, and covers domains such as marine biology, ecology, oceanography, and sea geology. Participants almost exclusively study the ecosystems of the Baltic Sea, its climate, and ecological issues concerning fishery, pollution and other human disturbances. Emma Beauxis-Aussalet attended the conference. She presented the project and the public query interface at the workshop on *Large data analysis in marine biology science: new possibilities through visual analytics*. She presented the user studies on user information needs. In particular, she discussed user needs for provenance information, and the uncertainty issues. Finally, she explained the interaction principles to explore the F4K data, and performed a live demonstration. She collected the following feedback:

- Reporting ground-truth based evaluations and the potential errors of video analysis software is necessary for adopting this new data collection technique.

- The Baltic Sea Science community deals with similar uncertainty issues in the usage of satellite imagery. This data collection technique is also novel, and has uncertainty issues. The collected data contain large levels of noise and needs specific computations to handle it. The methods to calibrate measurements, and deal with noise, are dependent on the ecosystem to study (e.g., the study of seabed geology or algae bloom need different methods). Some of these methods are still under-development.

- This community also deals with uncertainty in the models used to predict environmental conditions (e.g., wave height). Significant work is performed to evaluate error margins, the conditions that provoke errors, and solutions to deal with them.

- The proposed visualization of uncertainty was well received (e.g., Fig. 4,5). The simplification of complex measurements, such as ROC curves and Precision/Recall, was appreciated for allowing non-experts to access the technical information without overwhelming them.

- An expert in data visualization shared difficulties in dealing with user needs with regards to uncertainty issues. Issues are difficult to tackle because i) dealing with uncertainty is a complex task involving several factors of uncertainty, and several measurements of uncertainty; ii) users may not be able to express their needs w.r.t. uncertainty; iii) data visualization experts do not have the necessary domain expertise to anticipate user needs; and iv) user trust is difficult to evaluate. Data visualization designers rely on an iterative collaboration with a domain expert to elicit the requirements for dealing with uncertainty.

- A biologist wanted to reuse the visualization system for exploring his own data, which are not video analysis data.

- The Baltic Sea Science community may use software similar to the F4K tool for studying the community of species living in specific environments. Baltic waters may be often turbid. Thus cameras would need to be equipped with a light, and image quality may be low. Further, this community would need to study organisms other than fish (e.g., crabs, growth of mussels).

4 Conclusion

The Fish4Knowledge public query interface addresses user information needs for exploring the data extracted from the videos, and for checking the uncertainties that can impact the results of data analyses. For these purposes, we designed and implemented novel visualization techniques. The visualization of video analysis uncertainty is drawn from ground-truth based evaluation of video analysis errors. The visual representation is simplified and easier to understand, compared to visualizations used by image processing experts, while preserving the essential information. The visualization of video analysis data addresses a variety of user needs. It introduces interactions for specifying what the graph axes represent and widgets for specifying the dataset of interest and for providing an overview of the multi-dimensional data.

The public query interface was well-received by the marine scientists we interviewed. The functionalities offered were relevant and usable. Essential user needs are addressed, however, these are not yet sufficient for complete working system. Marine biologists mentioned specific functionalities of interest for their particular needs. The public query interface does not accommodate all of these, but it supports the preliminary exploration of data, prior to specialized analysis with dedicated tools.

The improvement of the public query interface will continue after the end of the F4K project[11]. In particular, further research of interest concerns the study of uncertainties other than errors of video analysis software (e.g, varying camera's field of view, duplicated individual, sampling size variation), methods other than ground-truth based evaluations (e.g., risk of confusing species), methods for visualizing potential biases (e.g., Fig. 17), methods for fish count normalization, user control of ground-truth samples (and its impact on user trust), and the usability study of our proposed uncertainty visualizations and multi-dimensional visualizations. We assume such research can benefit other use cases, such as research using other multimedia collections, or needing to explore other multi-dimensional and uncertain data.

References

[1] Computer vision coral ecology. `http://cvce.ucsd.edu/index.php`.

[2] Ecotoxicology. `http://en.wikipedia.org/wiki/Ecotoxicology`.

[3] E Beauxis-Aussalet, E Arslanova, L Hardman, and J van Ossenbruggen. A case study of trust issues in scientific video collections. In *Proceedings of the 2nd ACM international workshop on Multimedia Analysis for Ecological Data*. ACM, 2013.

[11] For instance, some functionalities are currently under development and may not be available before November 15th (Fig. 4, 5, 6, 13, 14).

[4] E Beauxis-Aussalet, E Arslanova, J van Ossenbruggen, and L Hardman. Fish4knowledge deliverable D2.4 Advanced User Interface and component-based evaluation.

[5] E Beauxis-Aussalet and L Hardman. Fish4knowledge deliverable D2.2 User Scenarios and Implementation Plan. `http://Fish4Knowledge.eu/DELIVERABLES/F4K_Del2-2_v3-9.pdf`.

[6] E Beauxis-Aussalet, J He, C Spampinato, B Boom, J van Ossenbruggen, and L Hardman. Fish4knowledge deliverable D2.3 Component-Based Prototypes and Evaluation Criteria. `http://groups.inf.ed.ac.uk/f4k/DELIVERABLES/F4KDel23.pdf`.

[7] E Beauxis-Aussalet, S Palazzo, G Nadarajan, E Arslanova, C Spampinato, and L Hardman. A video processing and data retrieval framework for fish population monitoring. In *Proceedings of the 2nd ACM international workshop on Multimedia Analysis for Ecological Data*. ACM, 2013.

[8] E Beauxis-Aussalet, J van Ossenbruggen, and L Hardman. Fish4knowledge deliverable D2.1 User Information Needs. `http://Fish4Knowledge.eu/DELIVERABLES/Del21.pdf`.

[9] E Beauxis-Aussalet, J van Ossenbruggen, and L Hardman. Fish4knowledge deliverable D2.3 - Appendix I - How UI features address biologists requirements. `http://groups.inf.ed.ac.uk/f4k/DELIVERABLES/D2.3-appendix.pdf`.

[10] Graham P Harris. Temporal and spatial scales in phytoplankton ecology. Mechanisms, methods, models, and management. *Canadian Journal of Fisheries and Aquatic Sciences*, 37(5):877–900, 1980.

