

Fish4Knowledge Deliverable D2.4

Advanced User Interface and component-based evaluation

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

Abstract: We specify the designs for the advanced user interface for the Fish4Knowledge project. A first interface prototype was developed based on the user requirements reported in Deliverables 2.1 and 2.3. This allows users to visualize trends of counts of fish species and make comparisons. Feedback on this initial prototype and other interface mockups motivated the improvements to the user interface designs described in this deliverable. The success of the Fish4Knowledge system as a whole depends on whether a visualized trend reflects a real underlying biological phenomenon or is an artifact of the system. While the system's interface is designed to support the basic user needs of selecting and visualizing data sets, a substantial part of the user interface design is targeted at supporting the user in understanding the various levels of processing and their impact on the trend being visualized. It is this aspect that we concentrate on in this deliverable and future user interface evaluation and development.

Deliverable due: Month 18

1 Executive Summary

This deliverable specifies the designs for the advanced user interface for the Fish4Knowledge project. Based on the user requirements discussed in Deliverables 2.1 and 2.3 a first interface prototype has been developed that provides a web-based interface to all data that has been processed and stored in the project's central data storage. Its interface allows users to select specific subsets of the data and visualize trends of counts of fish species in this subset over time, as well as comparisons among subsets. The implementation is a complete end-to-end system, providing users full access to all data generated by the fish detection and tracking components as well as the species recognition components.

Feedback on this initial prototype allowed improvements to the UI designs, described in this deliverable. It also made it clear that for every potential trend visualized, the first question is whether the trend reflects a real underlying biological phenomenon or is an artifact of the complete system. Answering this question is non trivial and requires a certain amount of understanding from the user of the strengths and weaknesses of the various components of the system, at a range of processing levels. As a result, a substantial part of the UI is targeted at supporting the user to understand the various levels of processing and their impact on the data being visualized. The rest of the UI is directly related to selecting and visualizing data sets.

To complement the lessons learned from the first prototype implementation, we conducted a user study among several marine biologists, focusing on their requirements for data provenance. That is, we studied how much marine biologists want or need to know about the computer vision analyses provided, to be able to use the results in their own scientific work. We discuss in what aspects the uncertainties in the computer vision analyses are comparable to the statistical uncertainties in data acquisition that are more common in marine biology. As part of this study, we developed alternative UI versions communicating data provenance information, with increasing levels of detail. The results of this user study are also reflected in the designs discussed.

Finally, we exposed marine biologists to paper prototypes of previous versions of the designs discussed in this deliverable, focussing on the core visualization functionalities. This has resulted in the set of refinements to the designs discussed here.

Contents

| | | |
|----------|---|-----------|
| 1 | Executive Summary | 2 |
| 2 | Evaluation of experimental prototypes | 4 |
| 2.1 | User requirements | 4 |
| 2.2 | Experimental user interfaces | 4 |
| 2.3 | User interface refinements | 9 |
| 3 | The advanced user interface | 10 |
| 3.1 | The <i>Video</i> tab | 12 |
| 3.2 | The <i>Video Analysis</i> tab | 13 |
| 3.3 | The <i>Raw Data</i> tab | 24 |
| 3.4 | The <i>Visualization</i> tab | 24 |
| 3.5 | The <i>Report</i> tab | 29 |
| 4 | End-to-end evaluation | 29 |
| 4.1 | Early evaluation | 29 |
| 4.2 | Advanced user interface evaluation plan | 31 |
| 5 | Conclusions | 32 |

2 Evaluation of experimental prototypes

We first summarize the user requirements stated in Deliverables 2.1 and 2.3 (section 2.1), then briefly describe the experimental user interfaces used to investigate how to address the requirements and the results of initial evaluations carried out by potential users (section 3). We present the design implications and the user interface refinements derived from these evaluations (section 2.3).

2.1 User requirements

As reported in the appendix of the deliverable D2.3, the user requirements are mainly expressed in the Deliverable 2.1, and are complemented by the Deliverable 2.3. Some of them were excluded from the scope of the UI because of the comparatively high implementation costs. We summarize the user needs we address into the 5 requirements defined below.

D2.1-A - Support the analysis of population dynamics: We must support the analysis of population dynamics by providing the following metrics: abundance, species composition, species richness. These metrics can be calculated on specific datasets that target fish with specific characteristics (e.g., species, behavior, location...) *These metrics refer to the functionalities specified in Section 3 of Appendix VI of D2.1.*

D2.1-B - Support browsing of videos of interest: We must support browsing of videos of interest, especially the videos that correspond to the datasets that are visualized. *This refers to the video search functionality specified in the Appendix VI of D2.1.*

D2.1-C - Support the identification and the correlation of trends: We must support the identification of trends and correlations of trends in specific datasets and metrics of interest. *This refers to the most important queries envisioned by biologists, e.g., as detailed in Section 5 of Appendix VI of D2.1. Most of these queries involve the identification and the correlation of trends.*

D2.3-A - Provide an overview of the uncertainty inherent to video analysis components: For each video analysis component, we must provide an overview of the potential errors that are inherent to their specific video analysis function. This overview must report on i) the ground truth dataset, its inherent errors due to involvement of human judgment; ii) the related ROC evaluation of the components; and iii) the certainty score profile of the components. *This refers to the requirements expressed in sections 3.1 and 3.2 of D2.3.*

D2.3-B - Provide an estimation of the potential errors contained in visualized datasets: We must provide an estimation of the potential errors that are likely to be contained in the specific datasets that are currently visualized. This estimation of errors is expressed using 2 types of inter-related metrics, i) the certainty scores, and ii) the estimated numbers of True Positives, False Positives and False Negatives. From a high-level point of view, this is meant for users to evaluate the level of confidence in the trends and correlations of trends they identify. *This refers to the requirements expressed in section 3.2 of D2.3.*

2.2 Experimental user interfaces

To refine and evaluate the functionalities and the user interface designed in D2.3, we developed 3 experimental user interfaces as described below. The first interface aimed at exploring the available video data, and at investigating the potential engineering issues. The second

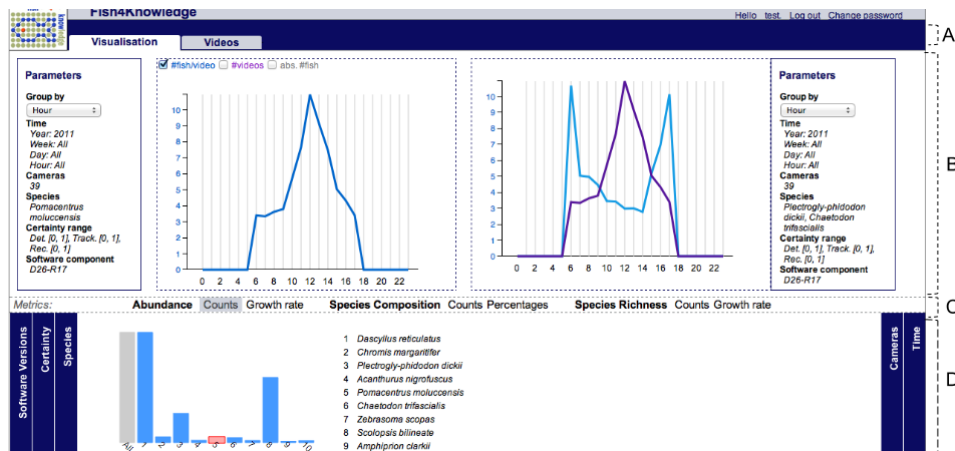


Figure 1: The first implementation of the *Visualization* tab.

interface aimed at refining user needs for data provenance information, and contributes to solving trust issues. The third interface aimed at testing the usability of the functionalities for data exploration and visualization (i.e., the *Visualization* tab as described in the section 4.3 of D2.3).

First implementation - We implemented a first experimental interface for collecting early user feedback, and for exploring the means to implement the web interface (see Fig.1). This interface implements a first version of the *Visualization* tab and the *filtering widgets*, as described in section 4.3 of D2.3. It also implements a means to overlay several fish counts in a comparative graph. It was commented on by many collaborators within the project, and particularly by Prof. Shao during the project meeting in December 2012. We collected valuable insights concerning the tool usability and potential technical issues.

Regarding usability, our main insight concerns the changes of magnitude between various fish counts that can be calculated for different species and/or locations, time periods. In some situations, the small histograms of the *filtering widgets* can be too small to be visible or clickable, while other histograms are reaching the maximum height of the display space. In other situations, overlaying several fish counts implies displaying counts that can be 10 times bigger or smaller one to another (or more), and displaying them on the same graph. The change of scale makes some variations in fish counts almost invisible (e.g., flattened lines). These issues can potentially be solved by using a logarithmic scale for the axes of the graphs (for the histograms of the *filtering widgets*, or for the main graph).

Regarding the overlay of several fish counts in a comparative graph, we identified an important issue regarding the x-axis and the comparison of fish counts of different time periods (e.g., comparing fish counts in March 2011 with March 2012). The prototype visualization uses only a linear axis for time periods. In some situation, the fish counts to compare are displayed apart on the left and right sides of the graph, which makes fine comparisons difficult. These issues can potentially be solved by modifying the visualization widget, so that each fish count is displayed with a specific scale on the axis of the graphs (i.e., the axes are also overlaid). But the axes may not be readable if too many fish counts are overlaid.

We also identified issues regarding the management of the overlaid fish counts. For instance, the first implementation provides no means to remove a fish count from the comparative graph, or to put a fish count in front, on top of the other ones (e.g., if there are occlusions). Further, the

color coding is not sufficient for users to remember which fish count is represented by which line. This can potentially be alleviated by indicating a title for each fish count, and displaying it when users rollover a fish count. We concluded that comparing patterns in video data requires advanced UI functionalities, that do not necessarily imply complex overlay of graphs. Thus our strategy to address the requirement D2.1-C consists of introducing simple functionalities to group and comment visualization, in a dedicated tab as described in section 3.5. We also include advanced functionalities for data visualization that provide alternatives to overlaid graphs, as described in section 3.4. For instance users can choose between overlaid graphs, stacked graph, error bars, or data points.

About the potential engineering issues, we identified that the time to load data is acceptable with the current amount of video data available. But we might need to improve query optimization when a more complete dataset will be available. We could provide an overview of the number of videos processed so far, and of those not yet processed. The number of missing videos is important, and further investigation concluded that video encoding errors make a substantial part of the video collection unusable. This point is essential to address trust issues, and the requirement D2.3-B.

Data provenance - We explored the means to address the requirements D2.3-A and D2.3-B. We developed alternative versions of the data provenance information, with increasing levels of detail, as shown in Fig. 2. We evaluated the interface with 18 biology experts using a semi-structured interview and an online experiment. We exposed them to the concepts of the ROC evaluation, and introduced the ROC information in 3-steps with an increasing level of complexity. The materials used for this experiment (i.e., the contents of this experimental interface and questionnaires) are reported in the Appendix D2.4-I. The detailed results of the experiment are described in Appendix D2.4-II. This appendix is released as a confidential document, since it concerns a research paper that is currently under double-blinded review. In brief, our findings are:

- 1) Understanding the ROC explanations is likely to demand substantial time and effort from most of the biologists, and requires suitable didactic materials. Biologists, however, are likely to invest the necessary effort, or to delegate this task to a trusted collaborator.
- 2) Biologists' traditional statistical methods are likely to be sufficient to cope with video analysis errors, even if we do not provide ROC explanations (e.g., no details about the numbers of TP, FP, TN, FN). But providing the ROC explanations helps biologists deciding the appropriate statistical methods to use.
- 3) The provenance information, in addition to the ROC explanations of video analysis components, needs to include a review of the characteristics of the sampling method and potential biases (e.g., is there an even number of samples over location, periods, or recognized species).

Data exploration and visualization - We developed a series of paper prototypes in order to investigate further the requirements D2.1-A, D2.1-C and D2.3-C, focussing on the functionalities of the *Visualization* tab. Heuristic evaluations were performed by separately interviewing 2 user interface experts and 1 marine biology expert. These experts are independent of the Fish4Knowledge (F4K) project.

The experts were exposed to a series of mockups presenting advanced functionalities for data exploration and visualization. These advanced functionalities aim at supporting a more flexible usage of the F4K video data, at including the visualization of statistics that address trust issues, while remaining intuitive for users. We experimented with an advanced version of the *Visualization* tab that introduces: i) the direct manipulation of the axes of the graphs; ii) the

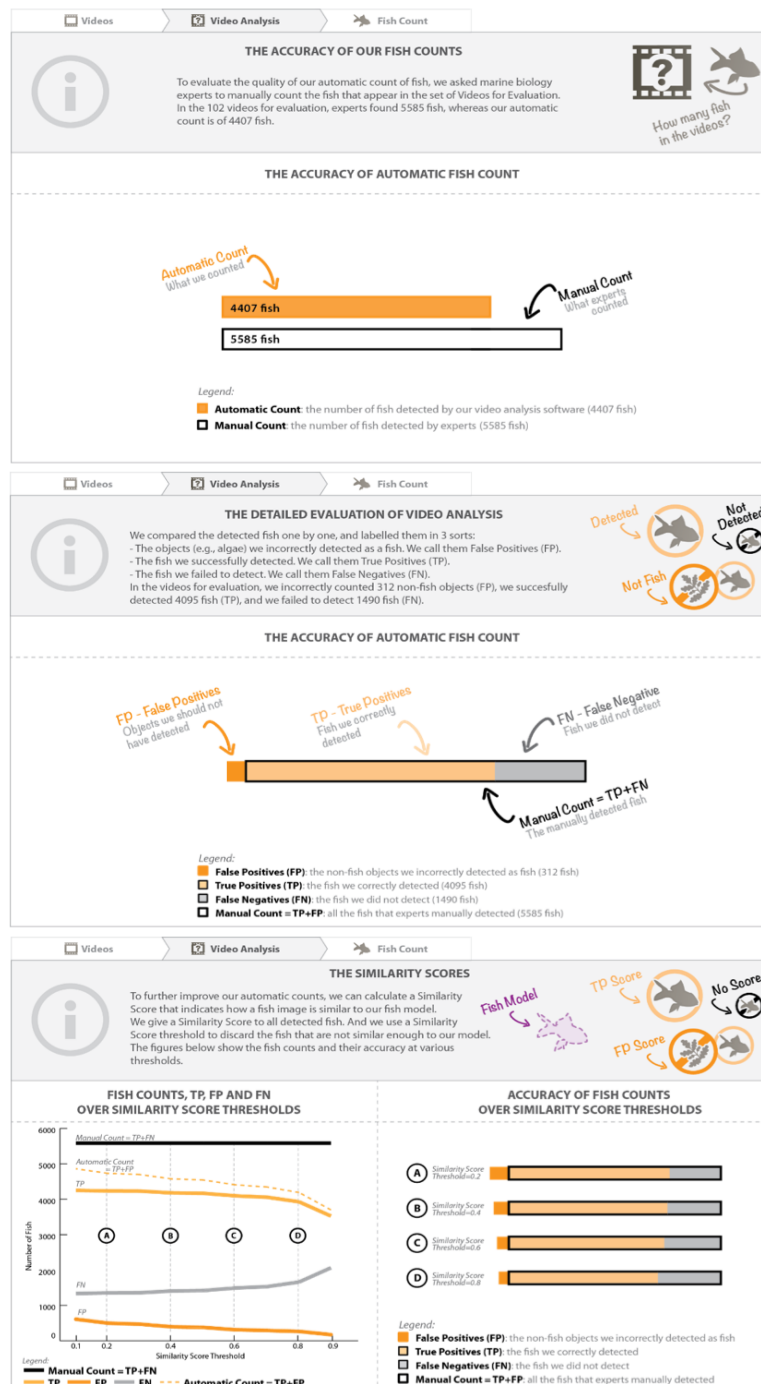


Figure 2: Three versions of ROC explanations for Fish Detection and Tracking with increasing levels of complexity (from top to bottom).



Figure 3: The most advanced version of the *Visualization* tab as evaluated by 2 UI experts, and by 1 marine biology expert.

decomposition the fish counts (or species counts) over 2 data dimensions (e.g., counts of fish per species and per month); iii) the introduction of functionalities to report and share users' insights in the F4K data. A complete report on the advanced *Visualization* tab is provided in section 3.4.

The UI experts explored the mockups in depth. They pointed out a series of flaws that mainly concerned the means to specify the dataset of interest, e.g., the *filtering widgets* to specify the periods and locations of interest (more details are provided in section 4.3 and Fig. 24 of D2.3). They were shown 4 versions of the *Visualization* tab, with increasing levels of complexity. They stated that the most complex version (shown in Fig. 3) is likely to be understood by users, and that there is less interest to experiment with the simplified UI alternatives we presented.

The marine biology expert did not explore the specification of the dataset of interest (e.g., the *filtering widgets*) in-depth. S/he, however, reviewed the functionalities for specifying the visualization of interest (i.e., the direct manipulation of the graph axes, the decomposition over 2 data dimensions), and the organization of the contents in the different tabs. S/he mentioned that the organization of content is clear, and is likely to cover all her/his information needs. S/he was shown the most complex alternative for the *Visualization* tab, which had been reviewed by UI experts (see Fig. 3). S/he quickly understood how to use the advanced functionalities. S/he did not show difficulties or hesitate when commenting on the tab. S/he indicated that the Raw Data tab is of crucial importance, as s/he would refer to it for deciding what specific visualization would be appropriate. S/he was enthusiastic when thinking about using such a tool, mainly because of the high flexibility allowed for visualizing data, and of the access to the underlying video processing details.

2.3 User interface refinements

Following the insights collected from the experimental prototypes, we elicited 4 main refinements of the UI presented in the deliverable D2.3. These refinements were used to design the advanced F4K user interface, as described in sections 3-3.5.

1. **Add functionalities to control the ground-truth collection.** Interviewed biologists were interested in exploring the ground-truth collection, as a means to control differences with the images in the general collection (changes of field of view, in fish abundance, time periods and areas covered in the ground-truth, differences in species composition between the ground-truth and the overall video collection).
2. **Modify the functionalities for data exploration and visualization:** the direct manipulation of the axes of the graph in the *Visualization* tab, described in section 3.4. These advanced functionalities allow a more flexible and open-ended use of the F4K data. For instance it allows the following data analyses targeting uncertainty issues: counts of ground-truth items (y-axis) per species, locations, or hour of day (x-axis); counts of processed videos (y-axis) per locations, hour of day, or week of year (x-axis). This addresses the requirements D2.3-A and D2.3-B.
3. **Delete the widget for statistical variability.** The functionality for studying statistical variability (i.e., mean and standard deviation, inter-quartile range) is still present in the UI but is available through different means of interaction. The *Decompose* function described in section 3.4 is now assuming this functionality. This new organization of the UI functionalities is meant to display only the filtering widgets in the lower zone of the interface, without mixing the functionalities for filtering with the functionalities for statistics. This allows a more consistent UI layout.
4. **Delete the *Data Processing* tab.** This tab was relevant for the previous version of the *Visualization* tab, especially to specify the detailed computation of species compositions or growth rate. This tab is no longer needed for the advanced UI, since the advanced *Visualization* tab already allow users to control of the parameters used to compute species composition and growth rate.
5. **Add a *Report* tab** for users to report and share their analyses of the F4K data. This addresses the requirements D2.1-A and D2.1-C, as well as the trust issues and the requirements D2.3-A and D2.3-C. Biologists can group and correlate video data of interest, and include high-level comments regarding their insights into fish populations. Grouping, correlating and commenting visualizations also supports users in applying and reporting their strategy to cope with video analysis uncertainty. Further, the *Report* tab is a simple solution to the issues encountered when overlaying several fish counts on the same graph, as for the 1st implementation of the interface shown in Fig. 1. For instance, each fish count is reported in separate visualizations, but with the appropriate scale, and with authoring functions (add, modify, delete fish counts).
6. **Replace the widget for error correction** as defined in section 4.1.2 of D2.3. The error correction as described in this deliverable needs further investigation to demonstrate its robustness to the following situation: the distributions of TN, TP, FN, and FP in the

ground truth may not be equal to these of the rest of the video data. Differences in these distributions impact the accuracy of the error correction mechanisms: in some situations applying the error correction method introduces more errors. This has no impact on our goal of communicating the implications of the ground truth based evaluations of the video analysis components for any specific query results.

3 The advanced user interface

These sections describe the main functionalities of the advanced interface. We first summarize the strategy used to organize the information into 5 tabs, and then we describe the content of each tab.

From our user studies, we identified 6 levels of video information processing and interpretation, as shown in Table 1. Each level addresses specific information needs with specific video data, and implies specific uncertainty issues. Each level uses the data processed by lower levels. The identification of these levels of information provides guidelines for the design of the video search tool. Specific layers of information extractions are involved in answering user information needs. When answering user information, the video search tool must provide information from the appropriate level(s) of information, and with the data provenance that addresses the uncertainty issues implied at each level.

The uncertainty issues related to the sampling method (requirement D2.3-B) concern mainly the collection of video images on level 1. The uncertainty issues related to the ROC evaluation (requirement D2.3-A) concern mainly levels 2 to 4. The means to cope with uncertainty issues depends on each specific user need. It mainly consists of visualizing and correlating several sets of relevant data, i.e., at levels 5 and 6. For instance, it can consist of gathering several fish counts each day at 11pm and comparing the mean and standard deviation over days and cameras.

These levels of information were used to derive and organize advanced functionalities supporting user information needs. For each level of information, we derived the information that must be conveyed and the functionalities to interact with the information. We organized these functionalities into the 5 following tabs. The mapping of functionalities into tabs was designed so as to group together consistent information.

1. *Video* tab - The 1st tab is similar as in D2.3. It addresses the 1st level of information in Table 1, and supports the overview of camera locations and settings, and the browsing of video clips and maintenance logs.
2. *Video Analysis* tab - The 2nd tab addresses levels 2-4 by supporting the control of the video processing components involved, the ROC evaluations, and the similarity score threshold used.
3. *Raw Data* tab - The 3rd tab provides an overview of the available video data and its properties. It helps users in identifying the information to select for their particular study.
4. *Visualization* tab - The 4th tab addresses the level 5 and supports user-defined visualization of video data, including visualizations that inform trust issues.

Table 1: Level of information for the exploration of the F4K video data.

| Level of Information | Video Data | Uncertainty Issue | Information accessed via tab | UI features |
|-----------------------------|---|--|---|--|
| 6. Overall Study | Aggregated visualizations | Validity of correlations. | <i>Report</i> tab | Group, comment and share visualizations. |
| 5. Visualization | Aggregated counts & growth rates over time, location, species, behaviour. | Statistical variability. Validity of correlations. | <i>Raw Data & Visualization</i> tabs. | Overview of available data and possible visualization. Specification of ad hoc visualizations. |
| 4. Fish Behaviour | Counts & growth rate of behaviours. | ROC evaluation. Different versions involved. | <i>Video Analysis & Visualization</i> tabs. | Control ROC evaluation, thresholds of classifiers, characteristics of component versions involved. |
| 3. Fish Species | Counts & growth rate of species, of fish per species. | ROC evaluation. Different versions involved. | | |
| 2. Fish Detection | Fish counts or growth rate, video excerpts of fish trajectories | ROC evaluation without TN. Different versions involved. | | |
| 1. Video images | Video clips. Counts of processed or flawed videos (<i>algae, turbidity, moving background, low lighting, encoding errors?</i>). | Missing videos. Low quality images. Changes in cameras settings. | <i>Video & Visualization</i> tabs | Control the sampled locations & periods. Check cameras setting. Watch video examples. |

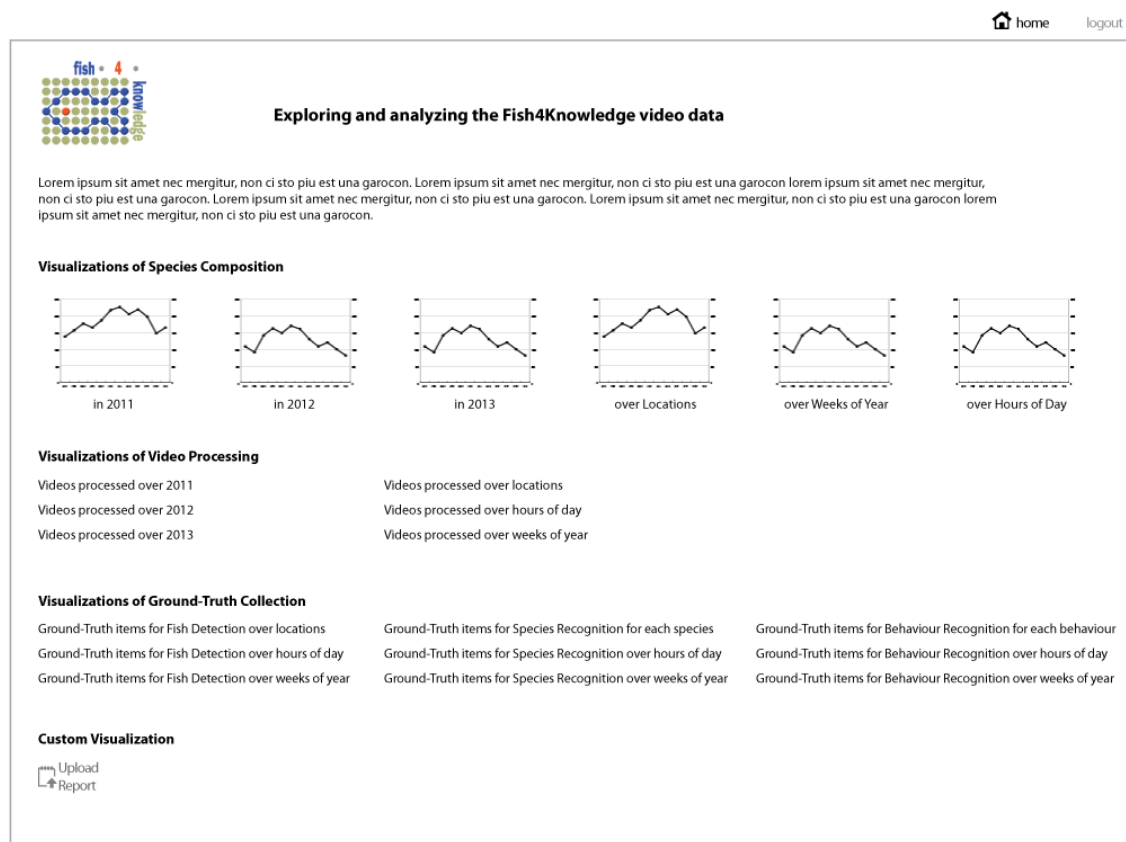


Figure 4: The home page with access to predefined visualizations.

5. *Report* tab - The last tab addresses level 6, and supports manual grouping and annotation of visualizations.

Additionally, we designed a *Home Page* that is displayed right after user login (shown in Fig. 4). It contains a short explanation intended for inexperienced users. It introduces a set of basic data visualizations that serve as a starting point either for getting familiar with the video data and the interface (i.e., for inexperienced users), and for beginning a new data analysis. Users can also upload a set of visualizations that were previously downloaded, and continue their analysis. In any of these cases, users are directed to the *Visualization* tab.

We intend to achieve a flexible design that can address a wide scope of potential usages. We particularly aim at supporting user-specified visualization of underlying video analysis processes, so that it allows the control of the potential biases introduced when using the video data for a particular scientific study.

3.1 The *Video* tab

The *Video* tab supports the requirements D2.1-B, by providing an advanced video browser, and D2.3-B, by providing access to maintenance logs and camera specifications. The content is organized in 3 sub-tabs. The 1st sub-tab (*Overview* in Fig. 5) provides an overview of the entire characteristics of the video collection: the location of camera, the video images, the important maintenance operations, the camera specification (e.g., resolution, lens, depth in the

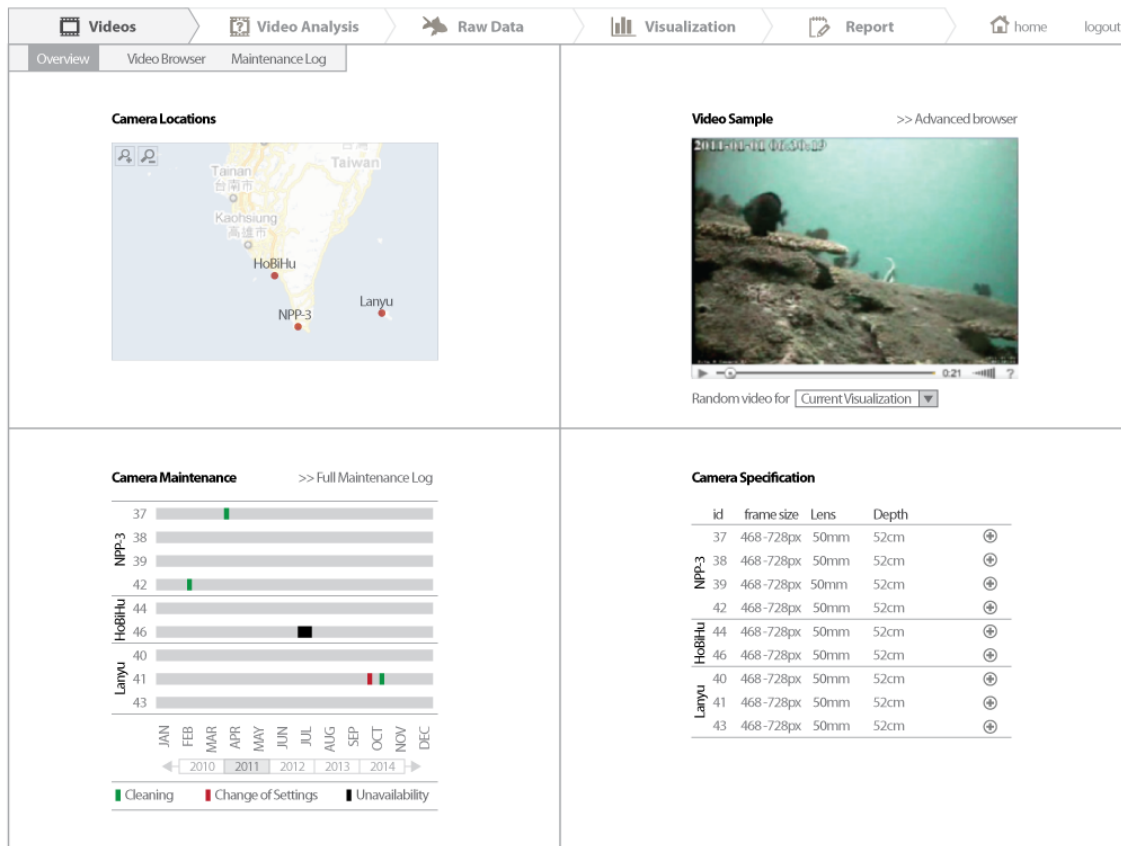


Figure 5: UI mockup for overviewing the characteristics of the video collection.

water). The 2nd sub-tab (*Video Browser* in Fig. 6) provides an advanced video browser that can retrieve videos containing specific species or behaviours. The last sub-tab (*Maintenance Log* in Fig. 7) provides both a tool to browse and retrieve maintenance operations, but also to edit the logs (e.g., add, delete, modify logs). We assume that biologists are likely to be aware of the maintenance operations, or other events happening to the cameras. We assume that they currently have no mean to record and centralize these events. This is why the UI supports the functionalities to add, delete and modify the maintenance operation logs.

3.2 The Video Analysis tab

The *Video* tab supports the requirements D2.3-A by providing advanced functionalities to explore the ROC evaluation of the F4K video analysis components. It also supports the requirement D2.3-B by providing the evaluation of the video analysis components that produced the video data currently visualized. It reports the ROC performance for the specific cameras, species and behaviours involved in the current visualization.

The content is organized in 8 sub-tabs. The 1st sub-tab (*Overview* in Fig. 8) provides an overview of the main characteristics the video analysis: the video processing steps (e.g., Fish Detection & Tracking, Species Recognition), the versions of algorithms applied over time, the ROC evaluation for each version, and examples of ground-truth items. It also contains a tool to request the specific analysis of a set of videos with specific versions of the components (i.e.,

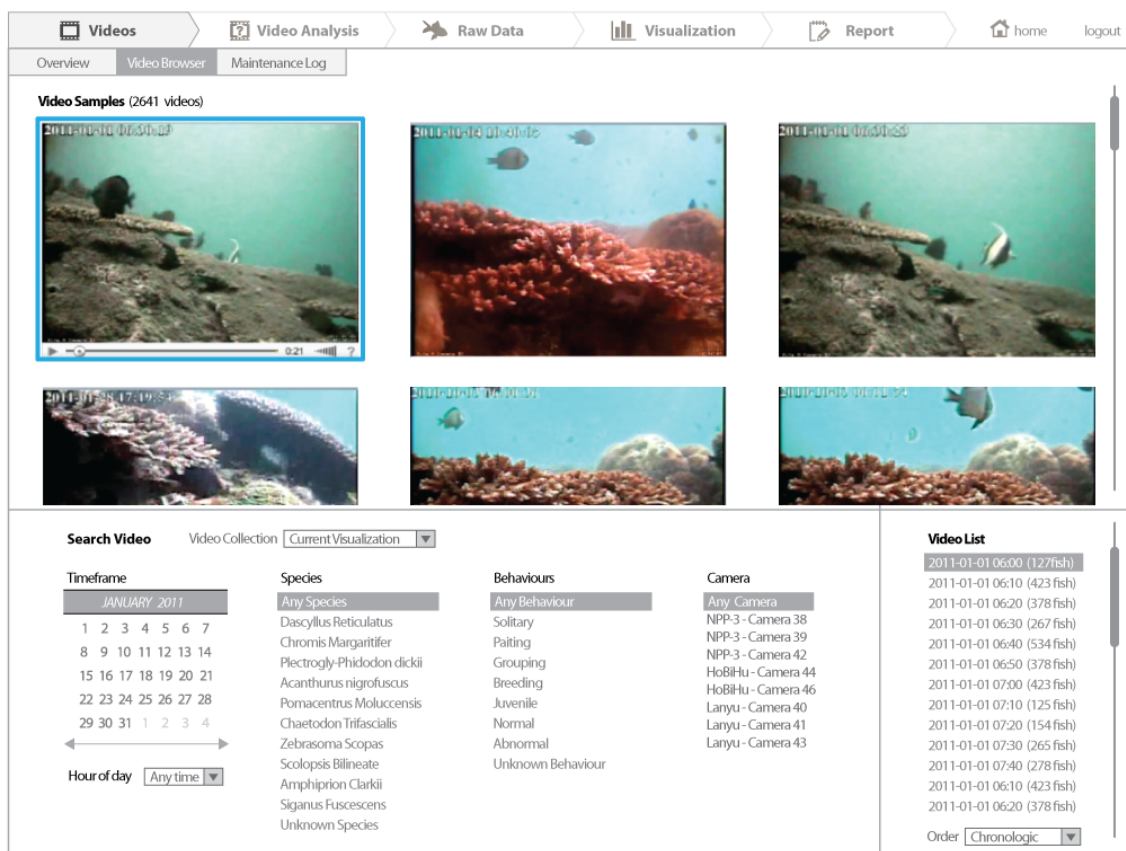


Figure 6: UI mockup for browsing the videos of the collection.

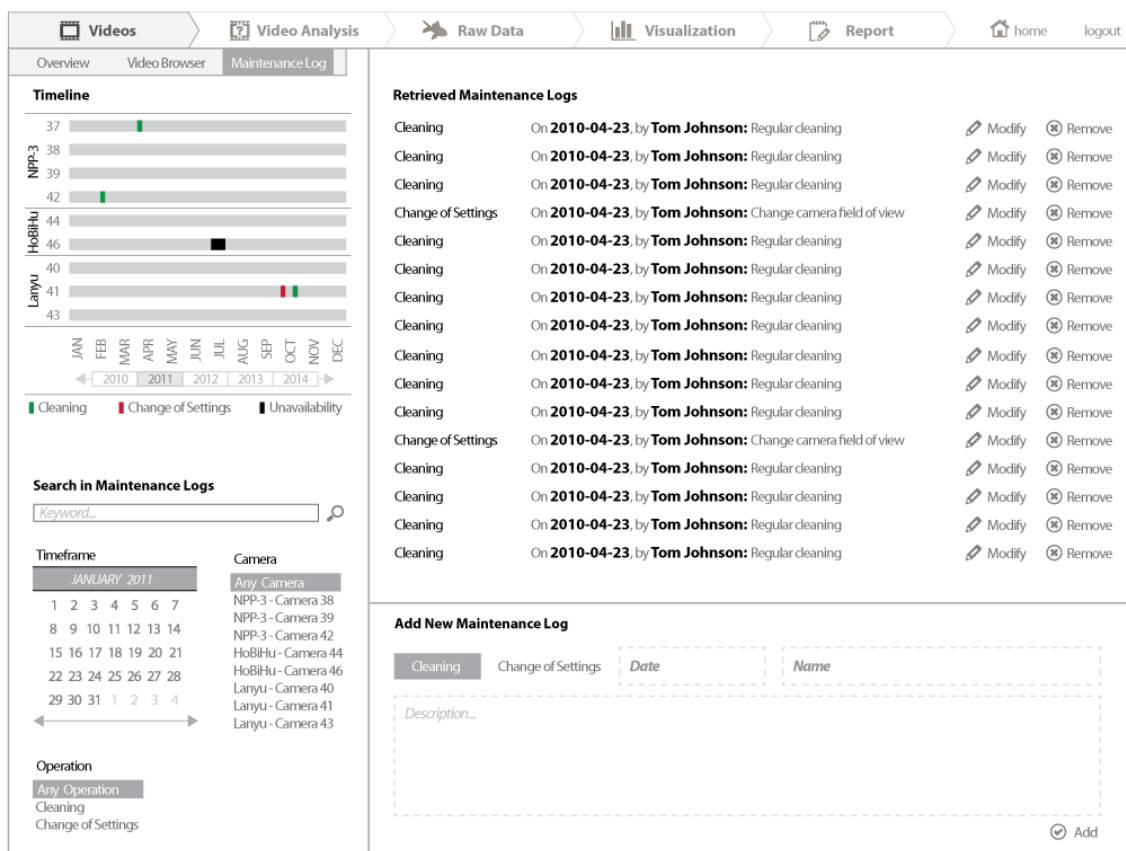


Figure 7: UI mockup browsing the reports of maintenance operations.

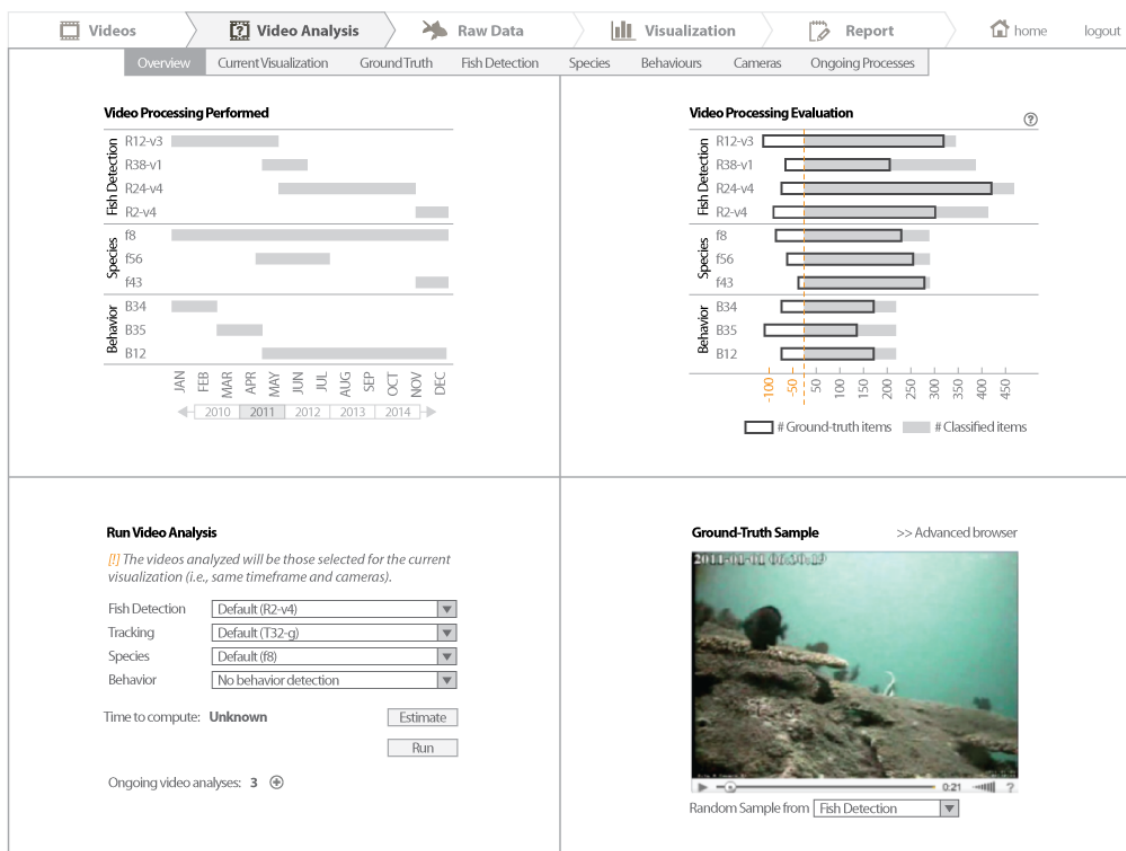


Figure 8: UI mockup for overviewing the video analysis components in use.



Figure 9: UI mockup for visualizing the video analysis characteristics for the current visualization.

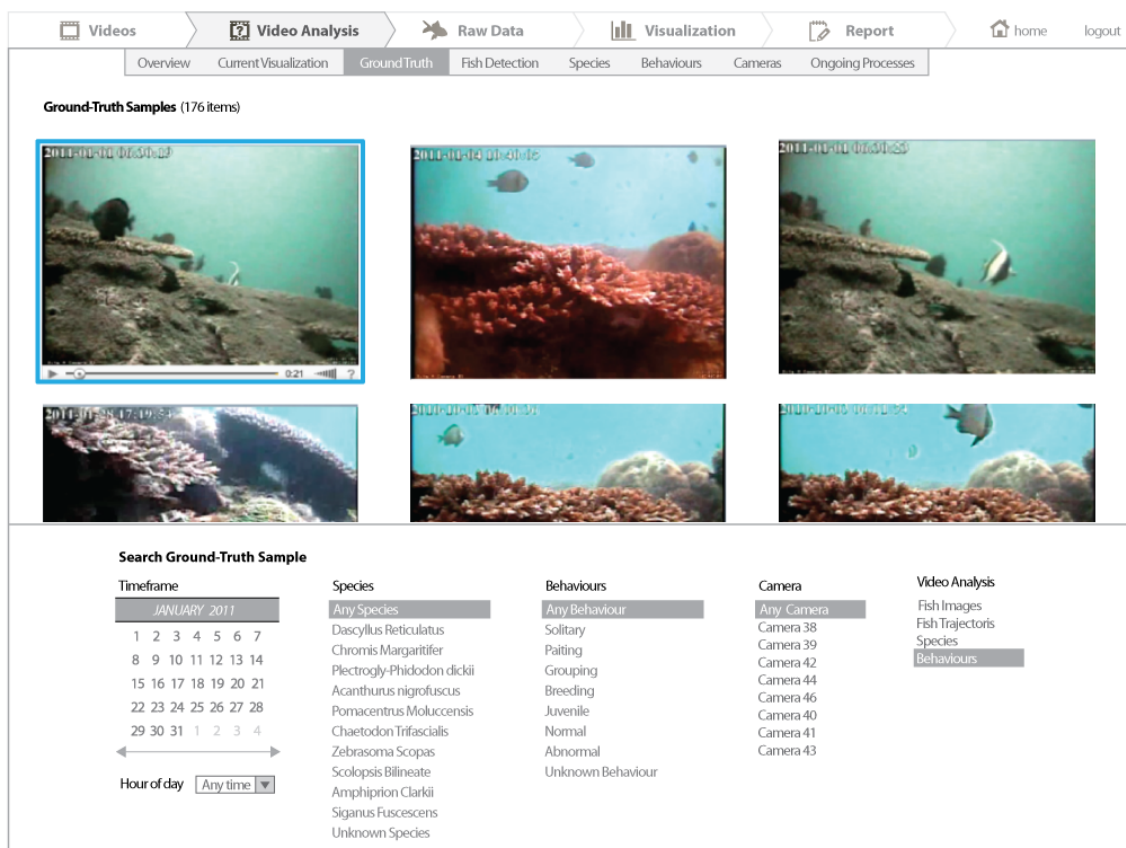


Figure 10: UI mockup for browsing the ground-truth collection.

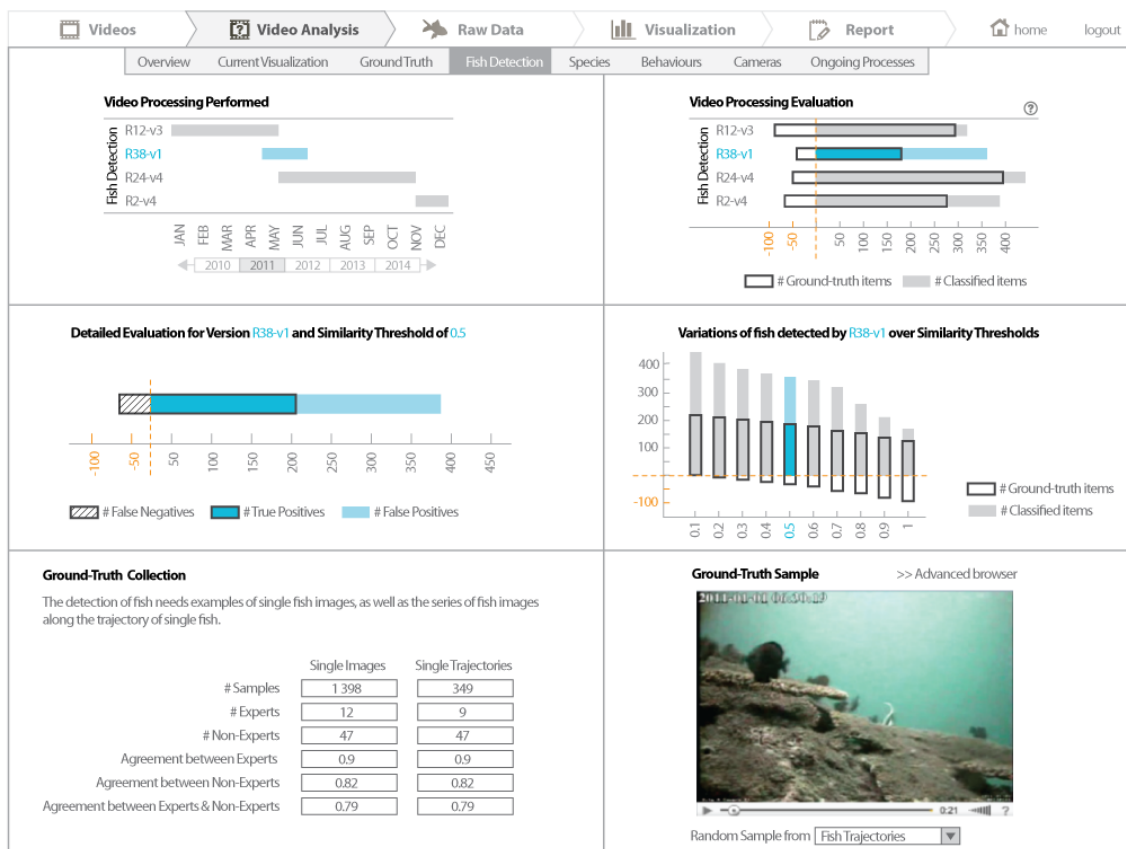


Figure 11: UI mockup for visualizing the characteristics of the Fish Detection and Tracking components.

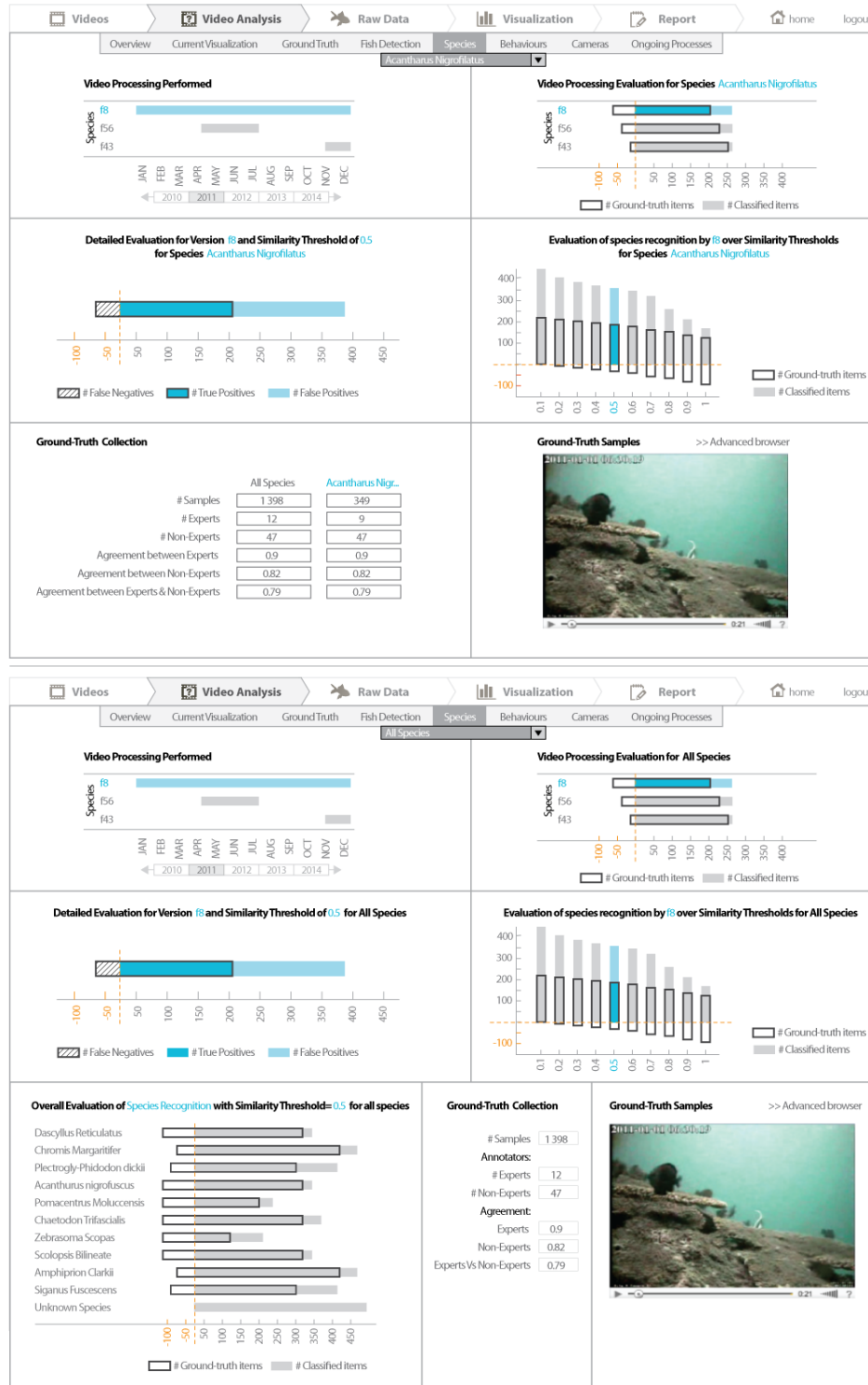


Figure 12: UI mockup for visualizing the characteristics of the Fish Recognition component.

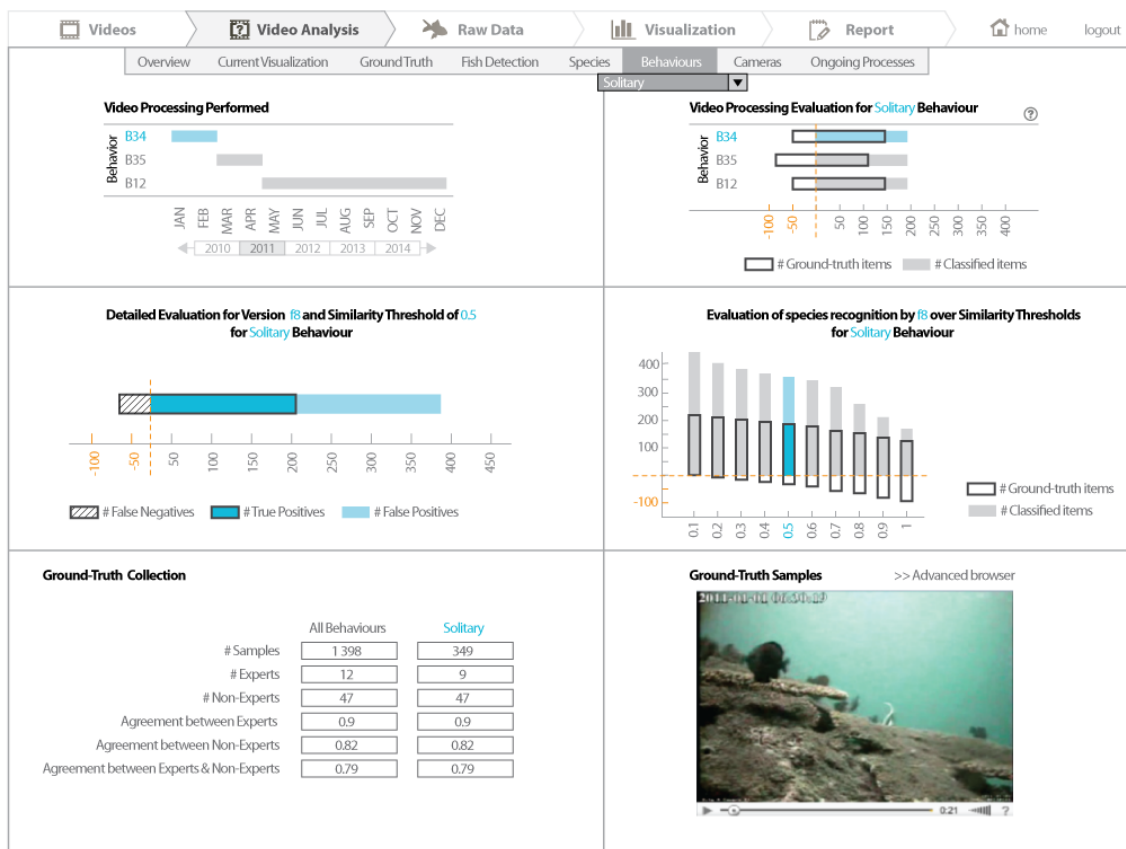


Figure 13: UI mockup for visualizing the characteristics of the Behaviour Recognition component.

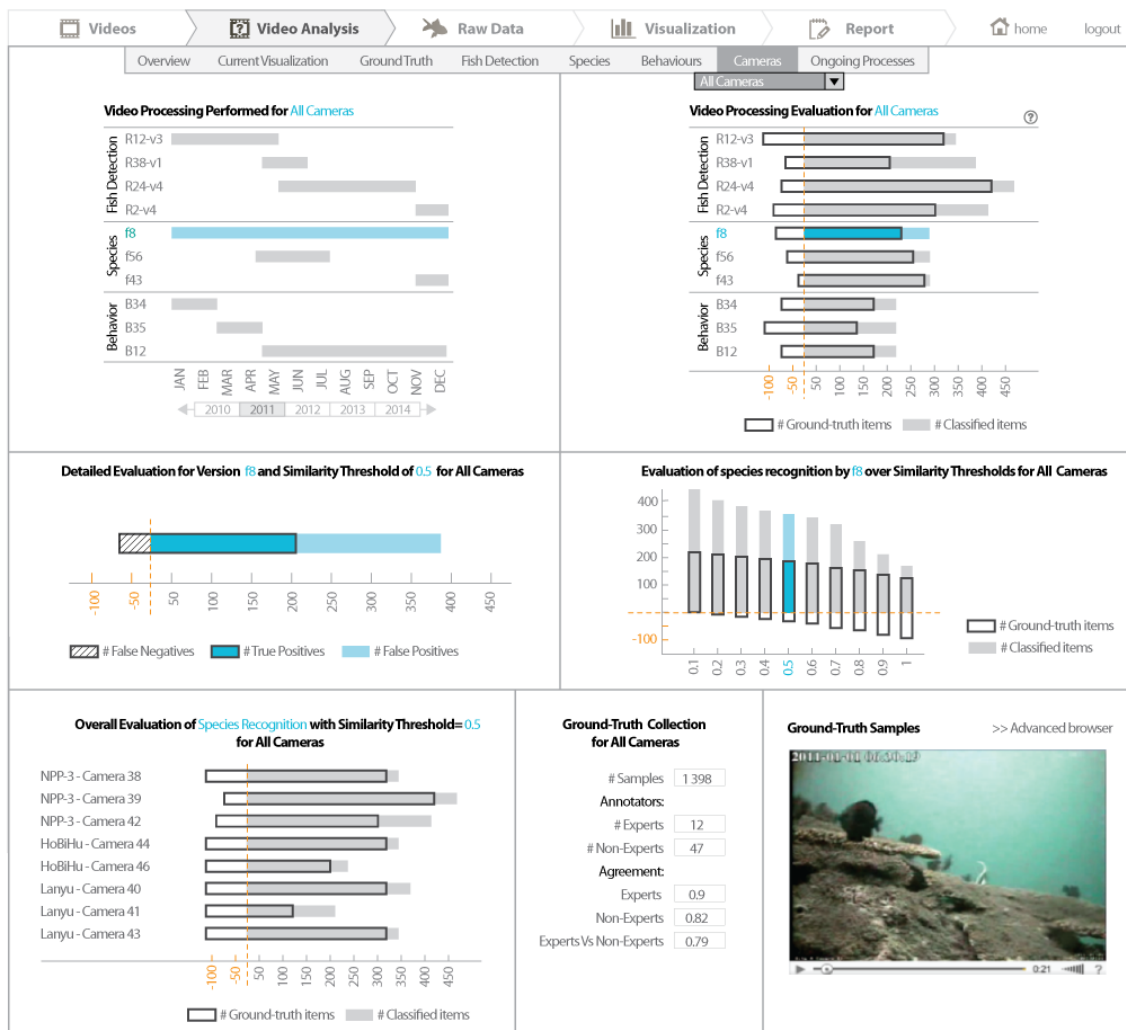


Figure 14: UI mockup for visualizing the video analysis characteristics for each camera.

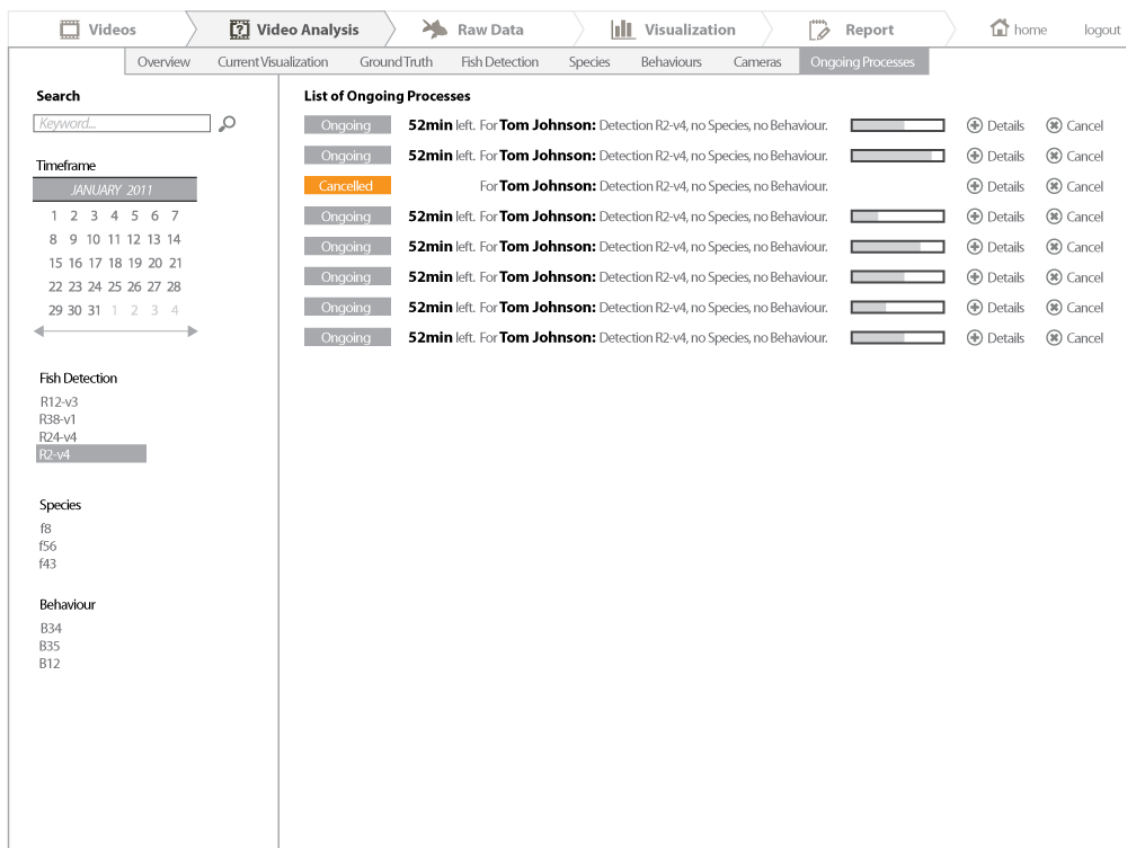


Figure 15: UI mockup for accessing on-demand video analysis processing.

advanced queries to the Workflow).

The sub-tab (*Current Visualization* in Fig. 9) provides the ROC evaluation of the components that produced the data currently visualized in the *Visualization* tab. It provides the detailed evaluation for each version of the components, and for each species and behaviours involved.

The sub-tab (*Ground-Truth* in Fig. 10) provides an advanced browser for retrieving specific items of the ground-truth (e.g., single fish images, examples of specific species, behaviours, period of time).

The sub-tab (*Fish Detection* in Fig. 11) provides ROC explanations for each version of the Fish Detection component. It also reports the number of items in the ground-truth, and gives some examples of these items. Similarly, the sub-tabs (*Species, and Behaviours* in Figs. 12 and 13) provide the ROC explanations for each version of the Fish Recognition and Behaviour Recognition components. They also provide the number of items in the ground-truth and some examples of these items. As shown in Fig. 12, users can visualize the evaluation for one specific species (or behaviour), or the overall evaluation taking into account all species (or behaviours). The sub-tab *Cameras* (Fig. 14) provides similar ROC explanations for each cameras.

The sub-tab (*Ongoing Processes* in Fig. 15) provides a tool to browse, retrieve and manage the video analysis tasks that were manually requested through the *Overview* sub-tab (the *special queries* to the Workflow).

3.3 The Raw Data tab

The *Raw Data* tab supports the requirements D2.1-A by providing an overview of the available data, and the characteristics that can be used for filtering. It also supports the requirement 2.3-B by including video data that can be used to evaluate the potential errors and uncertainties: the processed and erroneous videos, the versions of the video analysis components, the similarity score thresholds, and the number of items in the ground-truth.

The content is organized in 3 panels as shown in Fig. 16. The top panel provides a graph that represents the structure of the F4K data, from a user point of view. The boxes represent the entities that can be counted, or used to decompose other counts of entities (e.g., counts of fish over species). It also represents the characteristics of the data that can be filtered. The contours of boxes indicate the type of parameters used for filtering: *custom* filters with manually-defined parameter values, or *default* filters with automatically-defined parameter values. The default parameters concern the similarity thresholds (at EER point), the versions of the software components (as decided by the Workflow), and the species and behaviour (all of them).

The middle and bottom panels are the same as for the *Visualization* tab, and are described in section 3.4. They contain the functionalities for filtering the data (bottom panel), and for specifying the visualization of interest (middle panel).

3.4 The Visualization tab

The *Visualization* tab supports the requirements D2.1-A by providing an adaptable visualization system, with flexible means to filter the data and define what is represented by the graph. It contributes to addressing the requirement D2.1-C by supporting user-specified juxtaposition of small data visualizations (e.g., fish counts over locations juxtaposed to fish counts over time). It also supports the requirement 2.3-B by including video data that can be used to evaluate the potential errors and uncertainties: the counts of processed and flawed videos (e.g., presence of



Figure 16: The UI mockup of the *Raw Data* tab.



Figure 17: The UI mockup of the *Visualization* tab.

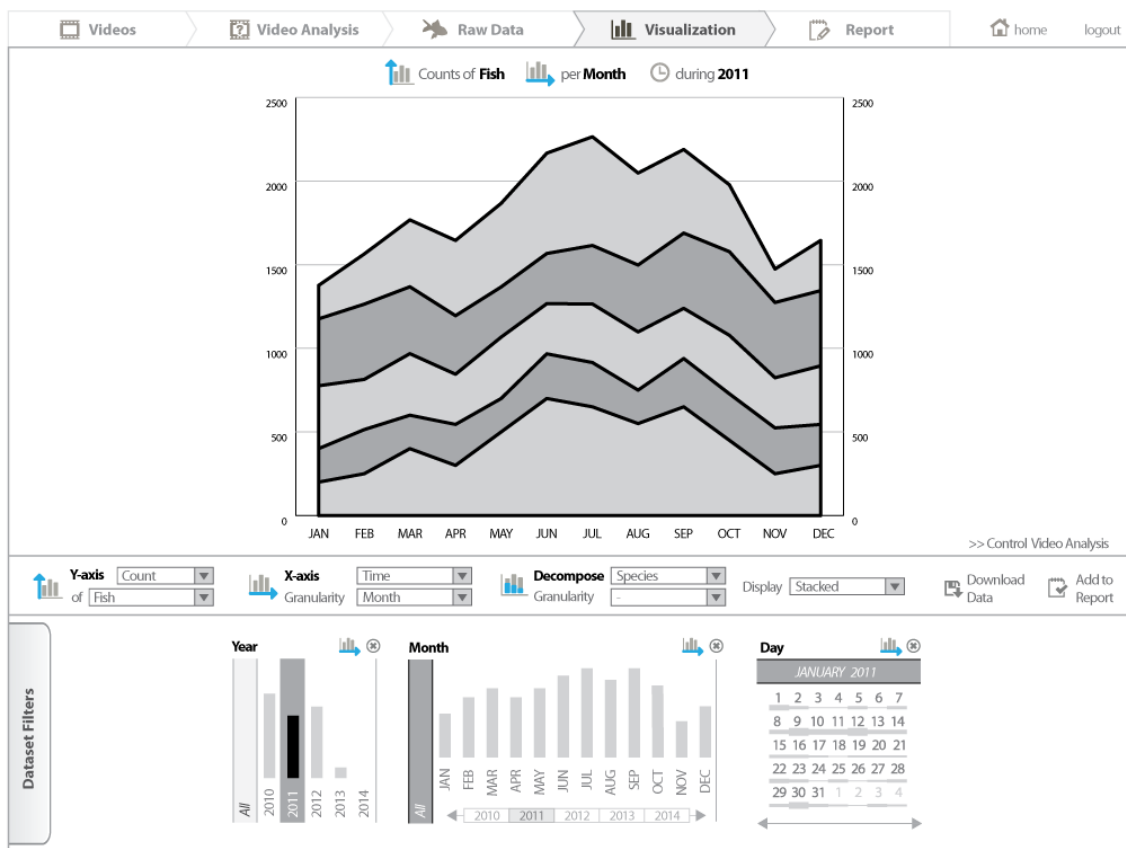


Figure 18: The UI mockup of the *Visualization* tab, with fish counts stacked for each species (usage of the *decomposition* function).

algae or encoding errors), the counts of the number of items in the ground-truth, the variations of counts (or growth rates) over the versions of the video analysis components, or over the similarity score thresholds.

The content is organized in 3 panels as shown in Fig. 17. The top panel provides a graphic that can be used to visualize counts (of fish, species, behaviours, videos, ground-truth items, or versions of components), or growth rates (of fish, species or behaviours). The counts and growth rates are represented on the y-axis. The x-axis can represent various dimensions of the data, depending on what users selected for the y- axis. For instance users can specify the following visualizations: fish counts on the y-axis and list of species on the x-axis; counts of species on the y-axis and week of the year on the x-axis; counts of processed videos on the y-axis and cameras on the x-axis.

Additionally, the graphic supports advanced functionalities to visualize data, and the *decomposition* of counts and growth rates. Such *decomposition* refers to the usage of basic statistical methods, as described in section 4.3 and Figs. 29-31 of D2.3. Considering the example of fish counts (y-axis) over species (x-axis), as illustrated in Fig. 17, we describe the usage of these advanced functionalities with the following **”decomposition” scenario**:

The user chooses to decompose the fish counts per species, and display them in a stacked graph as shown in Fig. 18. This allows biologists to study the species composition, and is basically the same functionality as the *Species Composition* functionality as described in D2.3. In the middle panel, the user interacts with the drop-down menus of the *Decompose* function. The 1st drop-down menu shows the dimensions of the data that can be used to decompose the fish counts. The users selects *Species*. There is no level of granularity for that dimension. But for instance, the *Time* dimension has levels of granularity such as year, month, week, day. In response to these interactions, the system calculates the fish counts for each species, and for each time unit of the x-axis. The 3rd drop-down menu lets users define what visual representation is required. The user selects *Stacked*. Alternatively, users can select *Error bar* to study the statistical variability. In this case, the system calculates the average fish count over species, and the standard deviation. The display of inter-quartile range is also available. In total, there are 5 types of graphical display available: error bars, inter-quartile ranges, single data points, stacked counts, and overlaid counts (e.g., similarly to the overlaid fish counts first implementation of the interface in Fig. 1).

The middle and bottom panels are the same as for the *Raw Data* tab. They contain the functionalities for filtering the data (bottom panel), and for specifying the visualization of interest (middle panel). The advanced functionalities for specifying the visualization of interest (middle panel) consist of the following:

- **Define the y-axis:** users can choose what is represented by the y-axis of the main graph. It can be either counts of growth rates. The y- axis can be counts of either: fish, species, behaviours, videos, ground-truth items, or versions of components; or growth rates of either: fish, species or behaviours.
- **Define the x-axis:** users can choose what is represented by the x-axis of the main graph. It can be either time periods, cameras, species, behaviours, similarity scores, versions of components. We also consider using bins of fish counts in the particular case where counts of species is selected for the y-axis, as the analysis of *species composition* performed by biologists often use this type of visualization.

- **Decompose the data:** users can break down the data over one extra dimension (as some form of z-axis). Each metric calculated for the y-axis (count or growth rate) over each bin of the x-axis is decomposed into a set of metrics calculated for smaller samples (see the “*decomposition*” scenario above).

The functionalities for filtering the data (bottom panel) are similar to those described in section 4.3 of D2.3, except that we discarded the widgets for *statistical variability* and for *error correction* (see section 2.3 for more explanations). Users can open a menu (the *Dataset Filter* retractable panel on the bottom left of Fig. 17) for launching the display of widgets that facilitates specific data filtering (e.g., for selecting the fish occurring only between 8am and 10am).

The advanced filtering widgets provide both i) the functionalities to set the filtering parameters (e.g., select only the fish from 1 specific species), and ii) the functionalities to juxtapose small data visualizations (e.g., fish counts over locations juxtaposed with fish counts over time). All the filtering widgets contain a small data visualization in the form of histograms. It represents the visualization that could be obtained in the data used for filtering was also used as the x-axis of the main graph of the top panel. We added a basic calendar for date selection. It provides overviews of the video data in the form similar to heat-maps, rather than histograms as the other widgets.

3.5 The Report tab

The *Report* tab supports the requirements D2.1-C by providing a simple mean to group and comment visualizations of interest. Users can add, delete, modify the visualizations, the titles and the comments. They can also download the reports, with the possibility to attach the raw data of the visualizations, and the evaluation of the underlying video analysis processes (e.g., ROC evaluation of the components involved). It consists of the same content as the sub-tab of the *Video Analysis* called *Current visualization*. It is illustrated by Fig. 9 and described in section 3.2.

Users can add visualizations in the report from the *Visualization* tab, by using a button on the right side of the middle panel. Reports can be saved and continued later by using the *Download Report* and *Upload Report* functionalities. Users can locally store a report, or share it with their colleague as a regular file (e.g., to be sent by mail). Reports can be uploaded and users can visualize or modify them. This offers basic but limited support to collaborative and long-term studies of the F4K data. It avoids the engineering problem of storing and sharing users’ data analyses online. For instance it would involve the implementation of a user right management system, for which we can not support the high implementation cost.

4 End-to-end evaluation

4.1 Early evaluation

Before implementing the advanced user interface, we will conduct a preliminary evaluation intended to identify major usability issues. This usability studies will be conducted with a limited number of marine biology experts (2-3 participants). They will be exposed to a click-through interface that mimics the behaviour of the advanced interface. We will develop static

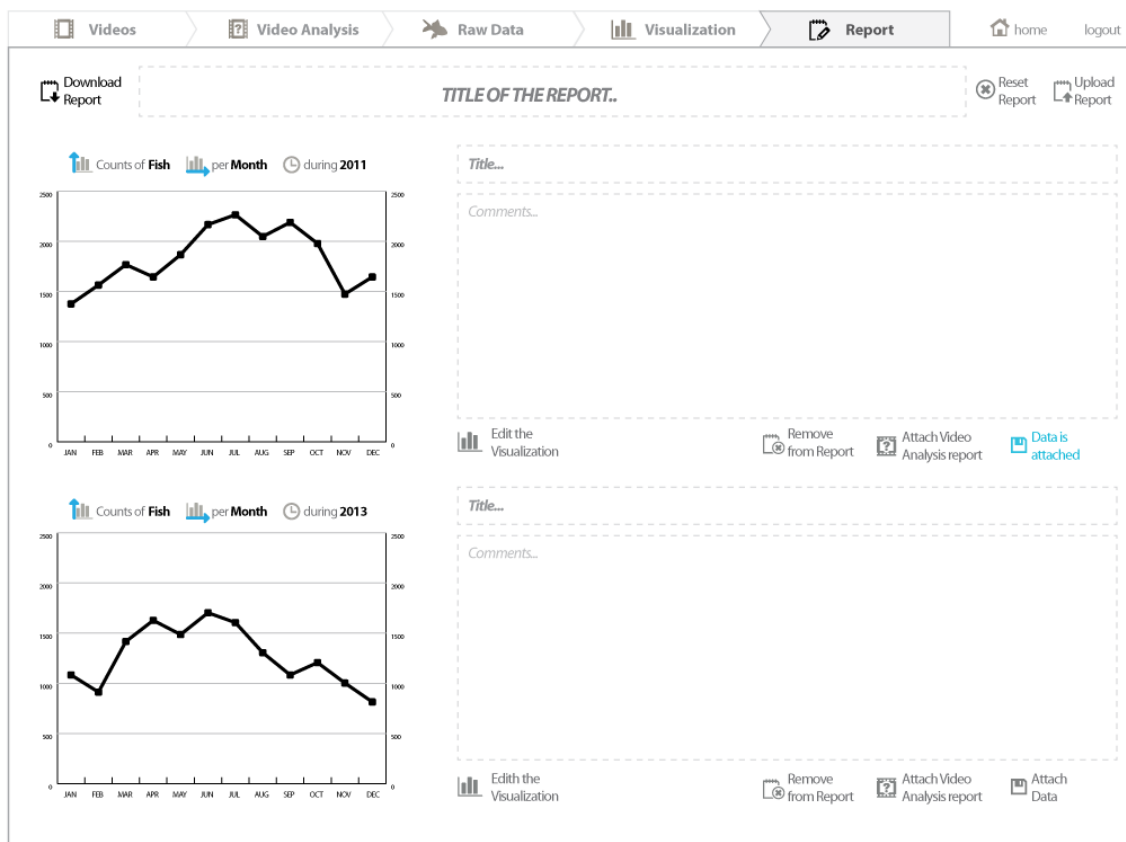


Figure 19: The UI mockup of the *Report* tab.

graphics and menus that only support a simplified data analysis task on a limited dataset (e.g., only fish counts over time, location and species are allowed). Participants will be asked to perform a data analysis task on that limited dataset, including the analysis of potential uncertainties. They will have to write a report on the *Report* tab.

The click-through interface uses the same graphical rendering as if it was implemented as a web interface. Thus the user experience should be similar as users should not notice a difference. The interface will present all the tabs described above. Only the possible visualizations will be limited, and the related data and graphs will be pre-computed. We will research the main usability issues, and particularly those that can be corrected before implementing the interface. Our main investigations are the following:

- *Home page*: Is the introductory text sufficient for lay users to understand the system? Are the default visualizations relevant? Do they need additional explanations?
- *Video* tab: Is the *Overview* sub-tab of interest for users? Is the quick video browsing in the *Overview* sub-tab of interest, or is the advanced video browsing sufficient (*Video Browser* sub-tab)? Are users likely to manually record maintenance logs?
- *Video Analysis* tab: Are the ROC evaluations understandable? Do we supply the right details regarding the video processing? Does the *Overview* sub-tab have similar issues as for the *Video* tab.
- *Raw Data* tab: Is the representation of the data understandable? Does it help user to understand what visualizations can be provided? Would user want to view the data of the current visualization as a table, i.e., similarly to the json or csv file that can be downloaded (using the *Download Data* button on the right side of the middle panel)?
- *Visualization* tab: Is the manipulation of the graph axes intuitive for users? Is it easy to decompose the data? Are the histograms of the filtering widgets interesting for users? Is it easy to use the mechanisms for data filtering? Do we need to consider the *average fish count per video* as a basic metric of the y-axis¹?
- *Report* tab: Is it easy to select and list visualizations of interest? Should we add functionalities to reorder the selected visualizations?

From this preliminary study, we will derive 2 sorts of insights on the interface: the high-priority modifications that will be implemented before the next evaluation (see section 4.2); and the design questions that require further investigations. These investigations will be performed during the next evaluation, and our findings will be applied to the final F4K user interface.

4.2 Advanced user interface evaluation plan

Our research questions build on those described in section 4.1 by including the complete F4K data and possible visualizations. This allows the evaluation of the advanced functionalities addressing the requirements D2.1-C, D2.3-A and D2.3-B. The evaluation of the advanced interface particularly focuses on the visualisation tab depicted in Fig. 17, and the minimum

¹Then if we decompose counts over, e.g., species, how do we do error bars? average fish counts per video have "native" error bars. Do we discard them when decomposing them?

interface to support users in a proper interpretation of the data shown in this tab. This leads to the following key research questions:

- To what extent are users able to use the selection mechanisms along the various dimensions to generate the visualisations of trends in fish counts over time?
- To what extent are users able to use the supporting tabs to decide whether a observed trend is caused by a system artifact or by a real underlying biological phenomenon?
- To what extent are users able to use the error metrics reported to justify the use or exclusion of automatic video analysis data in their own research?
- To what extent do users want or feel the need to watch (part of) the underlying video footage manually to confirm trends found by the automatic video analysis components?

We are currently implementing the advanced interface, as described in sections 3-3.5. This implementation will focus on the main visualisation tab and will be refined following the outcome of the early evaluation, as described in section 4.1. We plan to start the evaluation of the advanced interface in May 2013. We will recruit marine biology experts mainly from Taiwanese and Dutch institutions, for which we are in contact with around 20 researchers. We will also seek for the opinion of HCI experts that are independent from the project.

We will perform in-depth usability studies to evaluate how the user interface satisfies the requirements in section 2.1. Users will be asked to perform data analysis tasks as described in D2.3. The tasks include analyses of the potential biases that can be introduced through the video analysis processes. User will have to identify trends in the F4K data, correlate these trends, write a consistent analysis in the *Report* tab, and evaluate their level of trust in the outcome of their data analysis.

5 Conclusions

The initial end-to-end implementation of the basic UI functionalities, along with paper and click-through mockups of designs of other parts of the interface, were used as a basis for user evaluations. The evaluations were carried out with a number of marine biologists, who confirmed the need for access to provenance information in order to understand the trends they were seeing in the analysed video data. We also learned that in order to trust the results provided by the system they were less concerned by the simplicity of the explanation and more concerned by its accuracy.

These results, along with other observations made by these users, allowed us to create a series of improved designs for the advanced user interface. The implementation of these will be driven by the priorities of the provenance and trust issues of identifying trends in the data that we have.

The next set of evaluations is scheduled for May, starting with the meeting in Taiwan at the end of April. We will investigate specific provenance and trust questions, and also gauge the suitability of the overall interface for selecting appropriate subsets of the data and understanding the relation of the analysed videos to the reported counts, including the error metrics and the quality of the video footage.