TRUST ISSUES IN SCIENTIFIC VIDEO COLLECTIONS

Emmanuelle Beauxis-Aussalet, Elvira Arslanova,
Lynda Hardman, Jiyin He, Jacco van Ossenbruggen,
Tiziano Perrucci

USER INTERFACE CHALLENGES

- The F4K system provides a solution, but marine biologists are not yet aware of the problem
  - user tasks do not exist yet
  - a working prototype is needed to communicate with users
    - goals of project
    - potential for their own research
- Uncertainties in the data affects all scientific conclusions that can be derived from it
  - user understanding of their impact is critical
  - F4K data issues are different from data biologists have collected themselves

FOUR MAIN TOPICS OF STUDY

- Understanding and supporting users' information needs
- Understanding and supporting users' interpretation of uncertainty

USERS' INFORMATION NEEDS

Identifying information exploration tasks
and providing a user interface to support them

REQUIREMENTS IDENTIFICATION FROM EXPERT INTERVIEWS

What are the most important user needs?

- initial study (beginning of the project)
  - interviews with 3 marine biologists, distilled to “20 questions”
  - basis for first UI prototype
- follow up study
  - interviews with 11 Dutch marine biologists
  - focus on their data collection methods
  - provides context on how to communicate *our* data

DATA COLLECTION METHODS

Typical data collection methods

1. Baited **cameras**, stereoscopic vision
2. **Cameras** hovering the seabed
3. **Diving** with handheld cameras
4. Experimental **fishery**, with fish dissection
5. Commercial **fishery**, from the entire North Sea market
6. **Diving** observation, manual reports
7. **Cameras** on commercial vessels

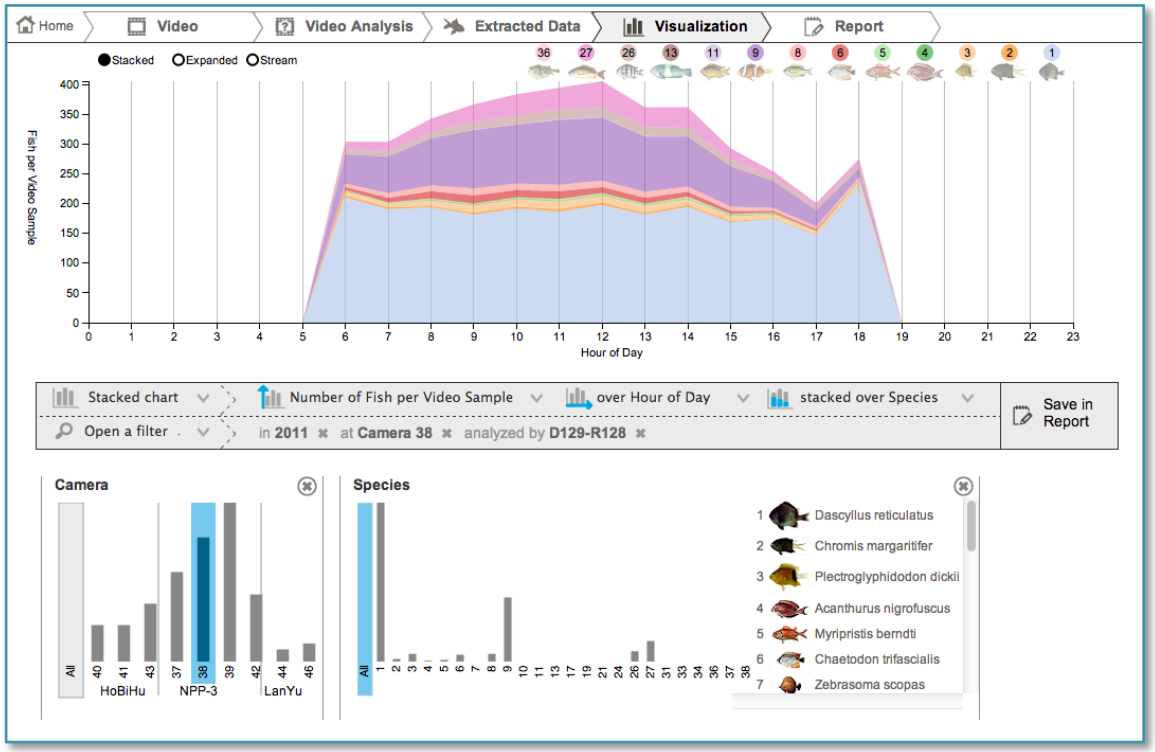
Uncertainty in current methods:

- Some species are very difficult to identify (e.g., camouflaged or hiding)
- Fish catches vary greatly under the same conditions
- Dense fish schools are difficult to count

USER REQUIREMENTS

HIGH-LEVEL, DOMAIN-ORIENTED INFORMATION

- **Abundance of fish** per species, location and time period
- **Number of species** per location and time period



EVALUATION OF USER INTERFACE BY EXPERT USERS

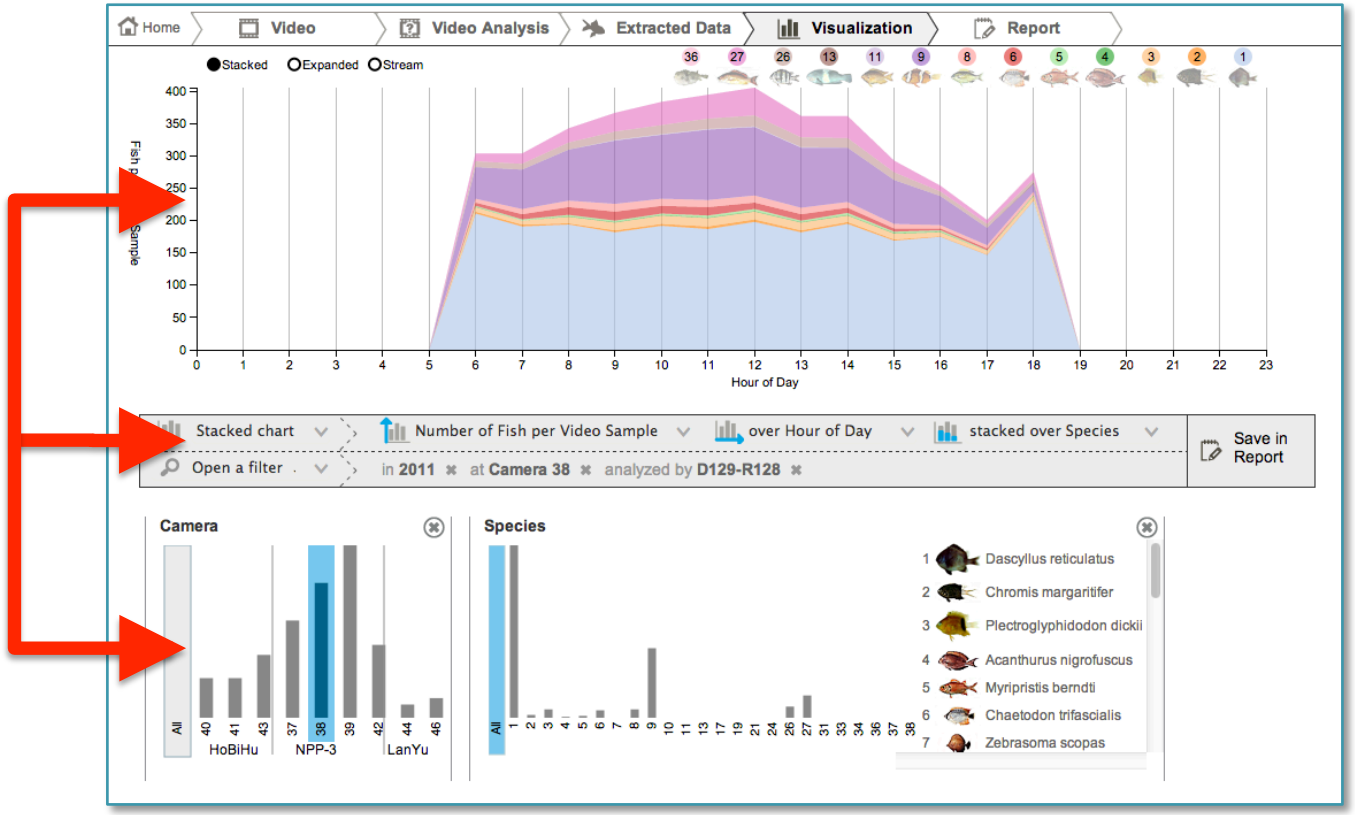
Ability of user interface to support representative information seeking tasks

STUDY SET-UP

- 10 participants (Taiwanese marine ecology)
- Representative tasks, such as
  - Is the abundance of species X less than species Y?
  - What is the count of species X at this time of year?
  - Does the same pattern occur in different locations?
- Levels of confidence measured by asking participants

RESULTS

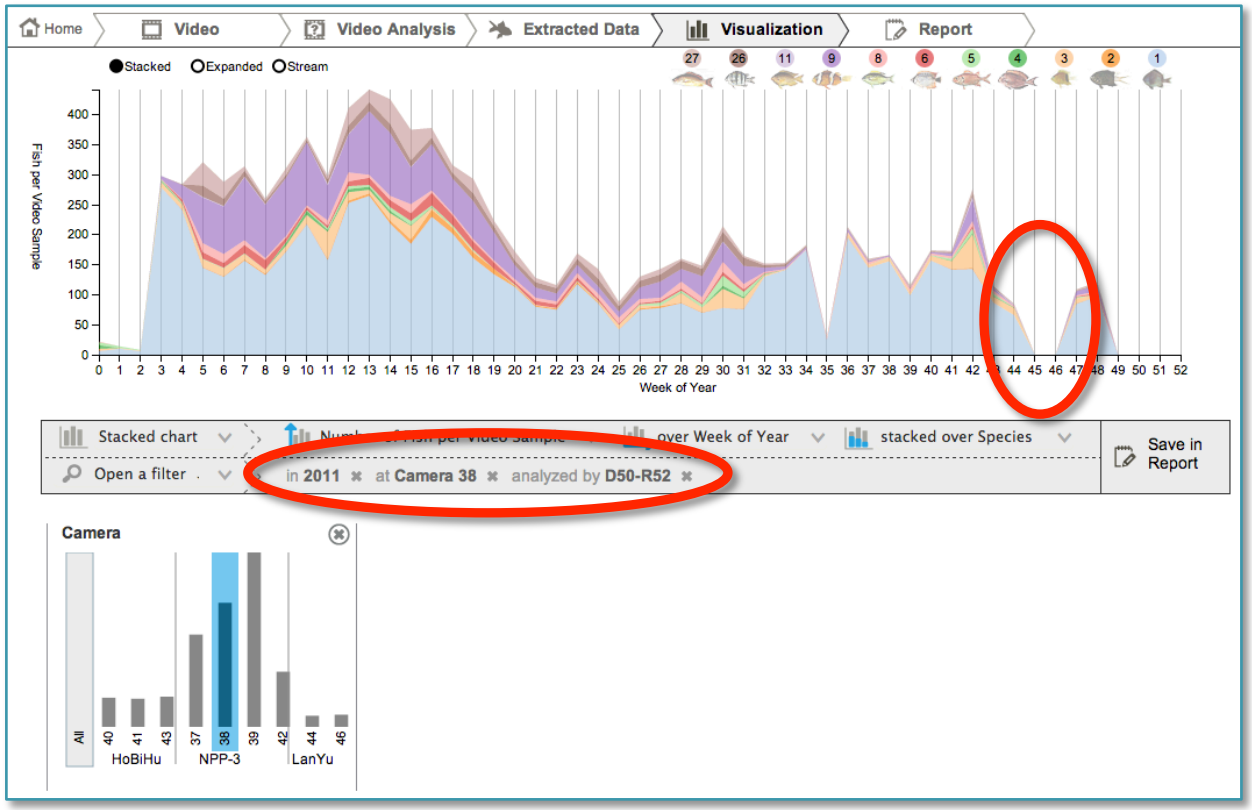
- USER INTERFACE -
EVALUATION



Interaction principles were easy to understand

Both the main graph and the filter histograms were perceived as useful

- USER INTERFACE -
EVALUATION



Numbers of videos and filters in use were not salient enough, or were overlooked by users

EVALUATION

Confidence	All Answers(%)		No Usability Issues (%)		With Usability Issues (%)	
	Right	Wrong	Right	Wrong	Right	Wrong
Very High	86 (43)	19 (9.5)	69 (**48.3**)	12 (8.4)	17(**29.8**)	7 (12.3)
High	55 (27.5)	10 (5)	40 (28)	8 (5.6)	14 (24.6)	3 (5.3)
Moderate	16 (8)	5 (2.5)	8 (5.6)	0 (**0**)	6 (10.5)	6 (**10.5**)
Low	1 (0.5)	6 (3)	1 (0.7)	4 (2.8)	0 (0)	3 (5.3)
Very Low	1 (0.5)	1 (0.5)	0 (0)	1 (0.7)	0 (0)	1 (1.8)
Total	159 (79.5%)	41 (20.5%)	118 (82.6 %)	25 (17.5 %)	37 (64.8 %)	20 (35.2 %)

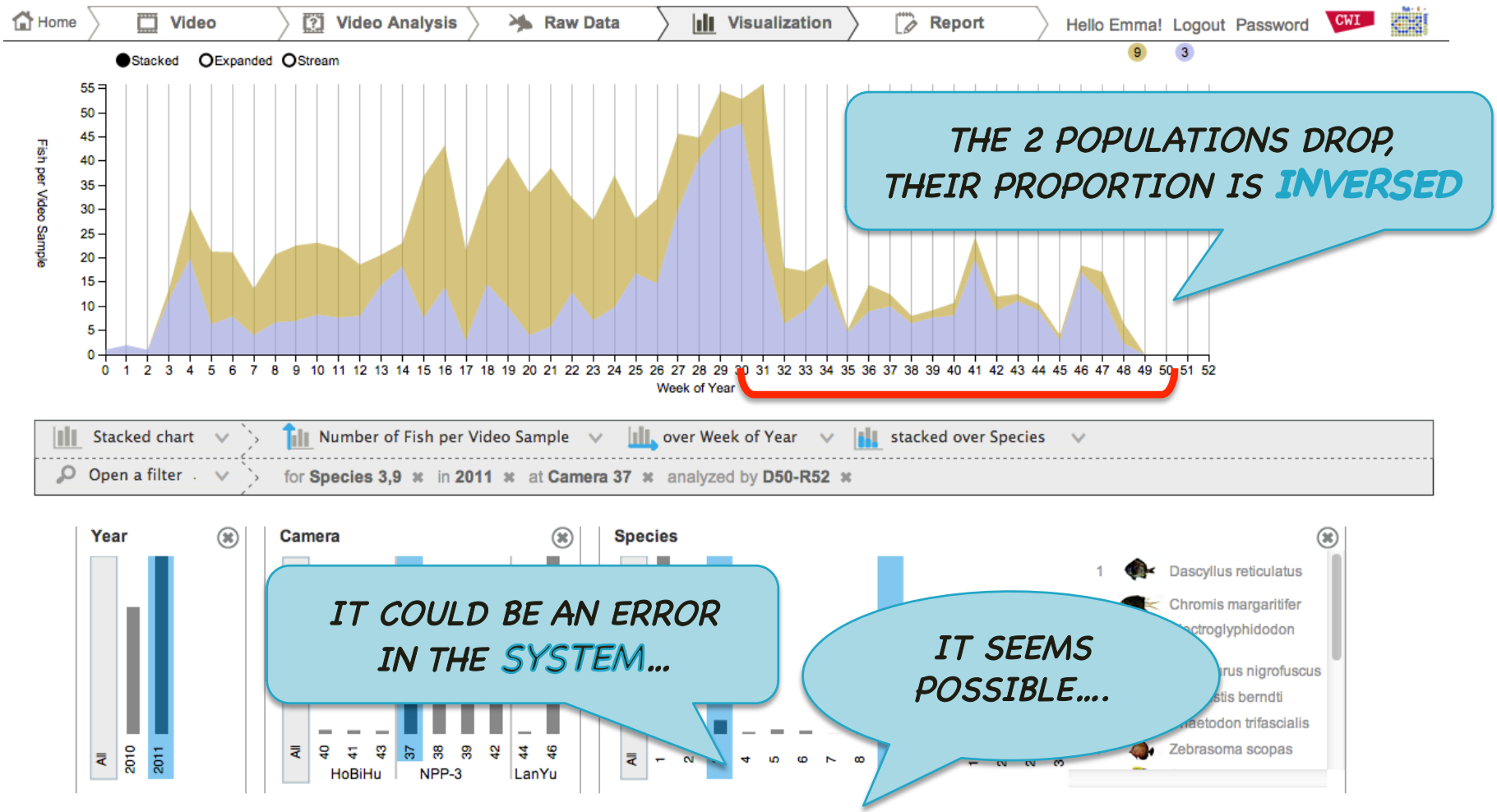
- **Users were mostly very confident in their answers, even when wrong**
- **Usability issues subjectively lowered user confidence**

INTERPRETATION OF UNCERTAINTY

Are users aware of the meaning and impact of uncertainty from video analysis

ARE THESE TRENDS SIGNIFICANT?

Do they represent the real fish community?



UNCERTAINTY FACTORS IMPACTING HIGH-LEVEL INTERPRETATION

Uncertainty due to video analysis techniques

- **Image processing errors**
- **Image quality**
  (blurred, algae, encoding errors…)
- **Noise and biases**
  (random and systematic errors)
- **Varying number of videos**
  (unprocessed or faulty videos)
- **Ground-Truth size and quality**

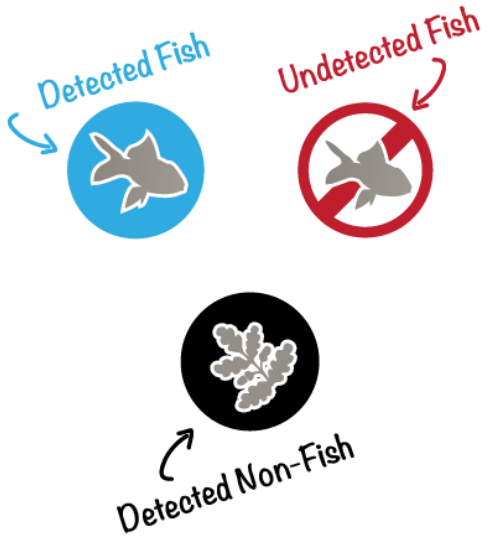
UNCERTAINTY FACTORS IMPACTING HIGH-LEVEL INTERPRETATION

Uncertainty due to the application context

- **Varying cameras' field of view** (the observed habitats impact the collected data)
- **Sampling with replacement** (individuals are repeatedly observed, depending on species' swimming behavior, and cameras' field of view)

EVALUATION OF IMAGE PROCESSING

What we investigated



- How to support user trust and acceptance of video analysis software?
- Can it be supported by reporting on ground-truth based evaluations?
- What technical details of these evaluations should be provided?

EXPERIMENTAL SET-UP

Explore responses to
ground-truth based evaluations

PARTICIPANTS & TASKS

Exploring practices across countries & data collection methods

11 participants with different cultures and expertise

- 3 levels of expertise: 2 Professors, 8 Researchers, 1 Master student
- 3 countries: Netherlands, Taiwan, Greece
- 4 data collection techniques: Video Images, Diving Observations, Experimental Fisheries, Commercial Fisheries

Tasks set-up

- Semi-structured interviews
- Explanations of ground-truth based evaluations
- Introducing technical details in 3 steps
- At each step, measurement of user trust, acceptance, understanding and information needs (7-scale and multi-choice questions)

3-STEPS EXPOSURE TO GROUND-TRUTH EVALUATIONS

Step 1 (left)

- Expert-made fish counts (i.e., from the ground-truth)
- Software-made fish counts

Step 2 (right)

- True Positives, False Positives and False Negatives
- No rates, just raw numbers

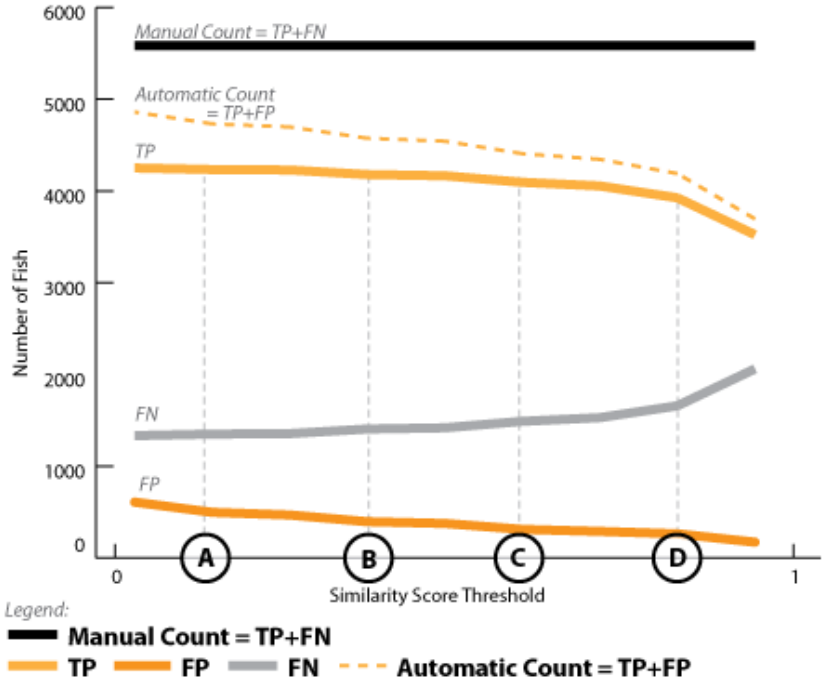
Step 3

- Similarity score threshold
- True Positives, False Positives and False Negatives, given for various thresholds
- Software-made fish counts, given for various thresholds

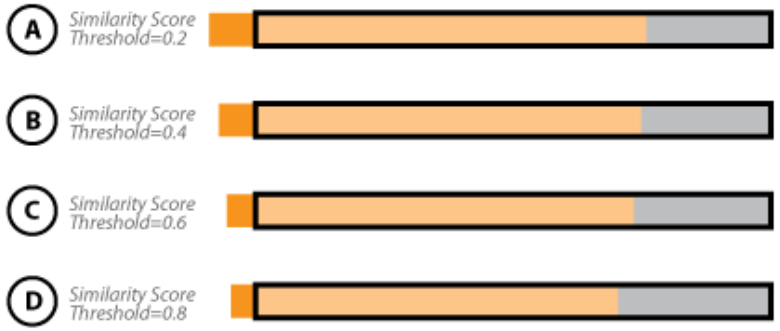
THE SIMILARITY SCORES

To further improve our automatic counts, we can calculate a Similarity Score that indicates how a fish image is similar to our fish model.
We give a Similarity Score to all detected fish. And we use a Similarity Score threshold to discard the fish that are not similar enough to our model.
The figures below show the fish counts and their accuracy at various thresholds.

FISH COUNTS, TP, FP AND FN OVER SIMILARITY SCORE THRESHOLDS



ACCURACY OF FISH COUNTS OVER SIMILARITY SCORE THRESHOLDS



Legend:

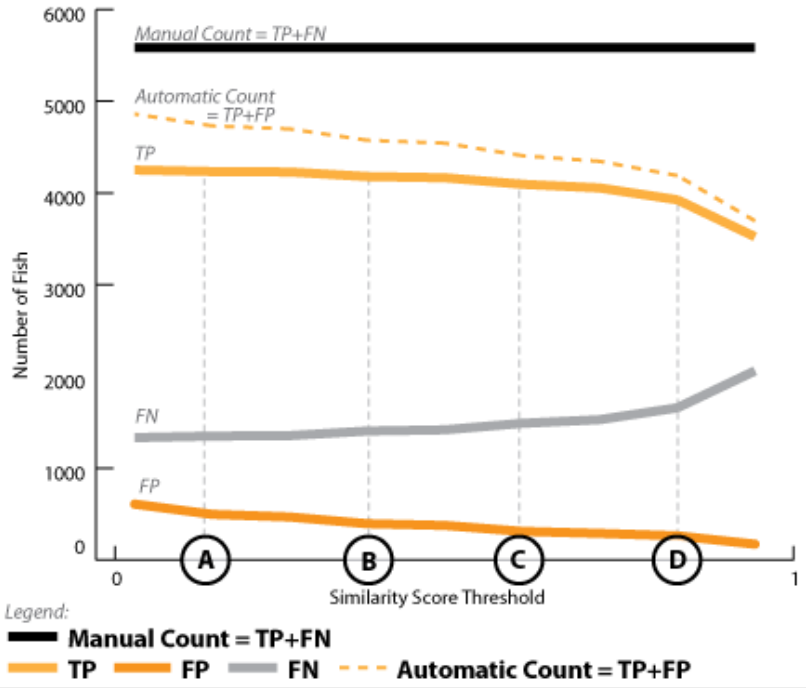
- **False Positives (FP)**: the non-fish objects we incorrectly detected as fish
- **True Positives (TP)**: the fish we correctly detected
- **False Negatives (FN)**: the fish we did not detect
- **Manual Count = TP+FP**: all the fish that experts manually detected

Videos | Video Analysis | Fish Counts

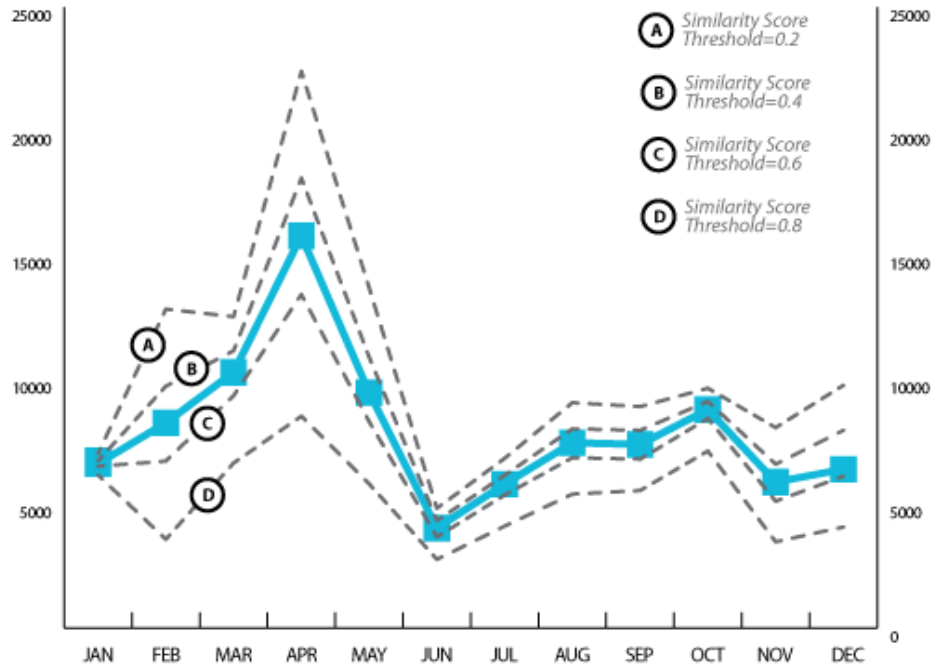
OUR COUNTS OF FISH

The counts of fish we detect can vary depending on the Similarity Score threshold we use. We used a Similarity Score threshold of 0.5. In the figure on the right, the dashed grey lines indicates the fish counts that can be obtained with different thresholds. The figure on the left shows what fish counts where obtained at these thresholds during the evaluation, i.e., when analysing the 102 videos for evaluation.

FISH COUNTS FOR EVALUATION OVER SIMILARITY SCORE THRESHOLDS



FISH COUNTS IN 2011 OVER SIMILARITY SCORE THRESHOLDS



RESULTS

Impact of ground-truth based evaluations on user trust and acceptance

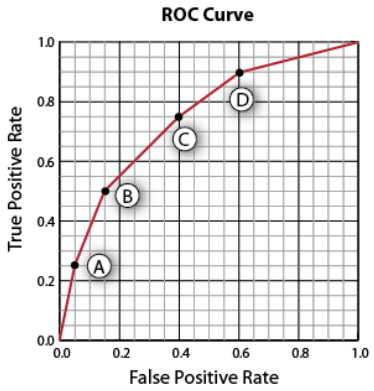
USER TRUST AND ACCEPTANCE

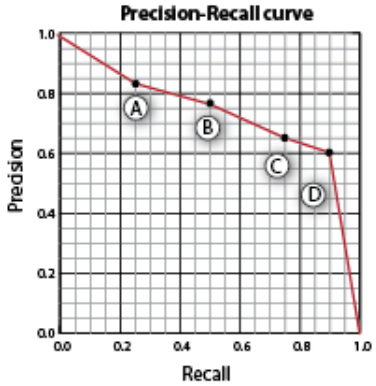
Potential support offered by ground-truth based evaluation

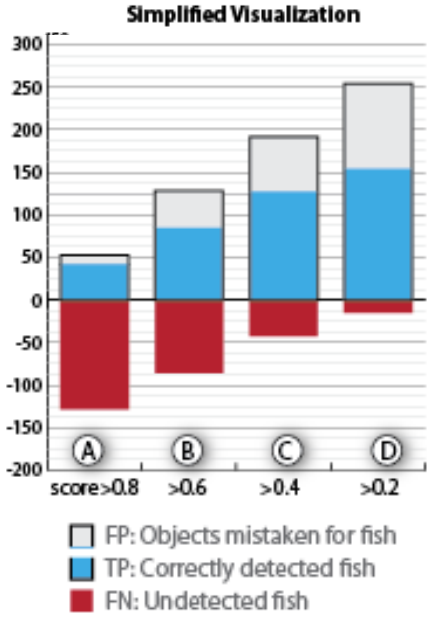
	Trust			Acceptance			Understanding			Information Need		
Case	Step 1	Step 2	Step 3	Step 1	Step 2	Step 3	Step 1	Step 2	Step 3	Step 1	Step 2	Step 3
2	+	+	-	+	+	+	-	-	-	++	++	++
3	-	-	+	+	+	+	+	-	+	++	+	+
4	++	++	++	++	++	++	+	- -	- -	+	+	++
6a	-	-	-	-	-	-	++	++	++	+	+	+
6b	- -	- -	-	+	+	++	++	++	++	++	++	++
7a	- -	- -	-	–	-	-	+	+	-	++	++	++
7b	+	-	+	+	+	++	- -	- -	-	++	-	+

- User **trust and acceptance** did not benefit from detailing ground-truth based evaluation
- **Understanding** was difficult, and additional **information needs** were not addressed
- This may cause the ground-truth based evaluations to be ineffective
- **Acceptance** remains relatively high, since automatic video analysis saves a lot of effort

DESIGN RECOMMENDATION FOR NON-EXPERTS







Relatively complex "ROC" evaluations

- require effort to understand
- remove user's attention from underlying information task

Alternative design takes results of study into account

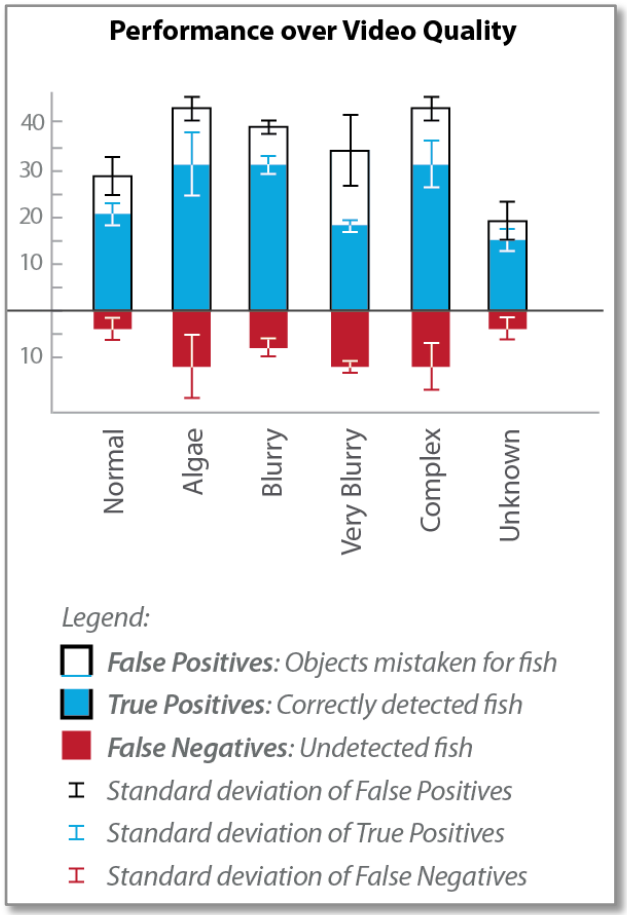
- reducing visual complexity

LIFE AFTER FISH4KNOWLEDGE?

FUTURE WORK ON UNCERTAINTY-RELATED INVESTIGATIONS

Development of methods for evaluating biases

- **Repeated ground-truth evaluation** under conditions that may induce biases, and compare means and deviations.
- Investigate alternative distributions of certainty scores, for example by applying **logistic regression** techniques to normalize fish counts.



CONCLUSIONS

- We established **information needs for scientific usage of data based on video analysis techniques**
- We delivered a **multi-dimensional data exploration interface**, addressing a large range of use cases, while being extensible to ongoing developments
- We provided a UI design to **bridge the knowledge gap** between video analysis experts and marine biology experts to address uncertainty issues

- USER INTERFACE -

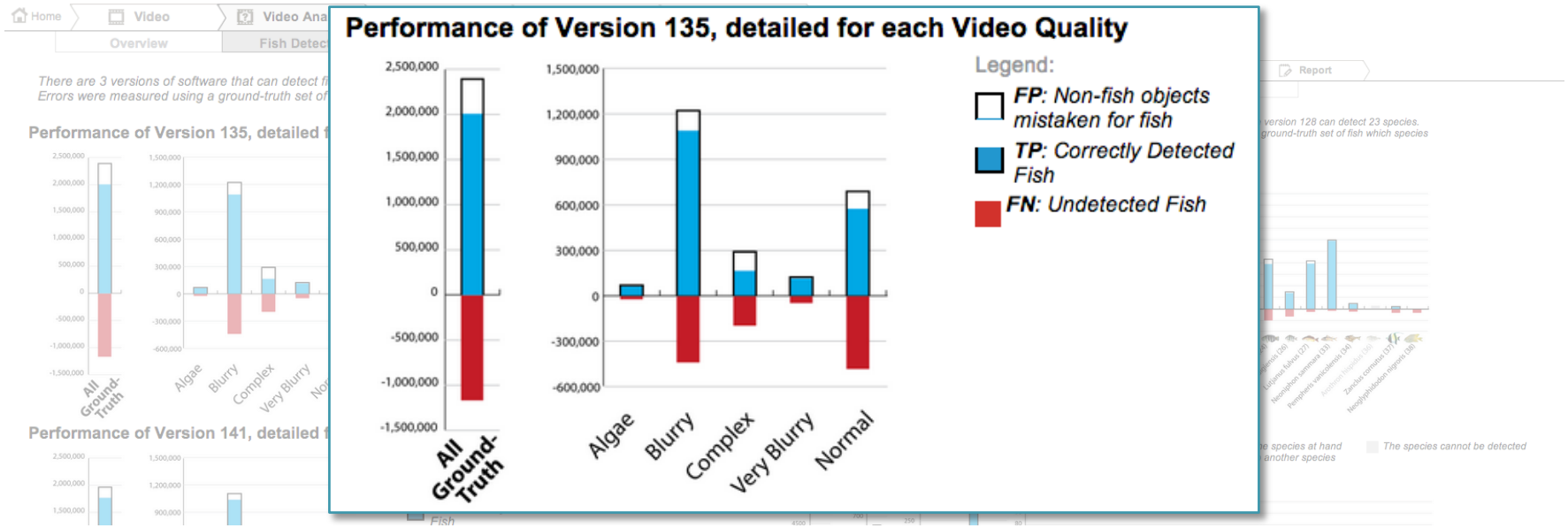
SUPPORT FOR UNCERTAINTY-RELATED INVESTIGATIONS

…and reports video analysis accuracy

for specific location, time periods, image quality

- USER INTERFACE -

SUPPORT FOR UNCERTAINTY-RELATED INVESTIGATIONS

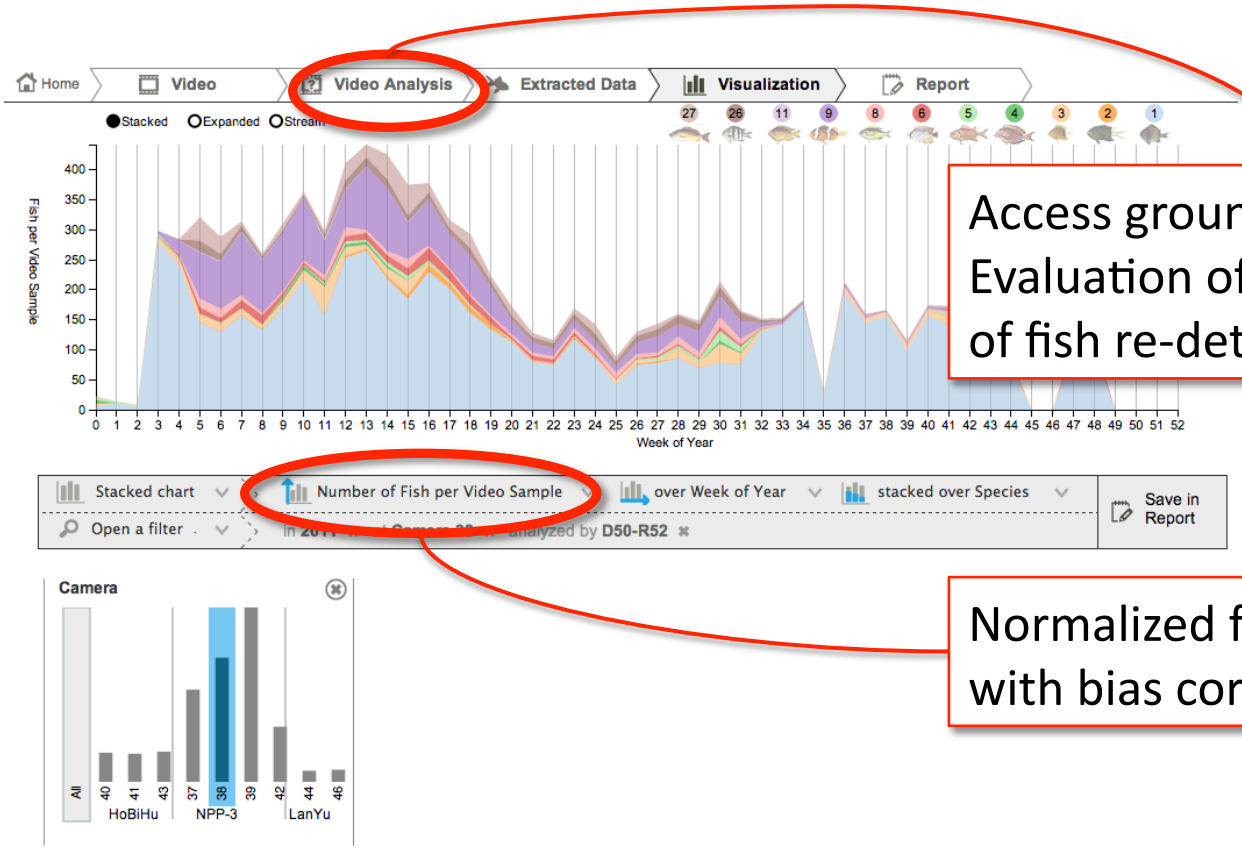


We studied ground-truth evaluation for non-experts

- Understanding the technical concepts demands important effort
- Other uncertainty factors also require user attention
- Simplified presentation of ground-truth evaluation are desirable

We propose a design excluding rates and True Negative

USER INTERFACE
FUTURE WORK

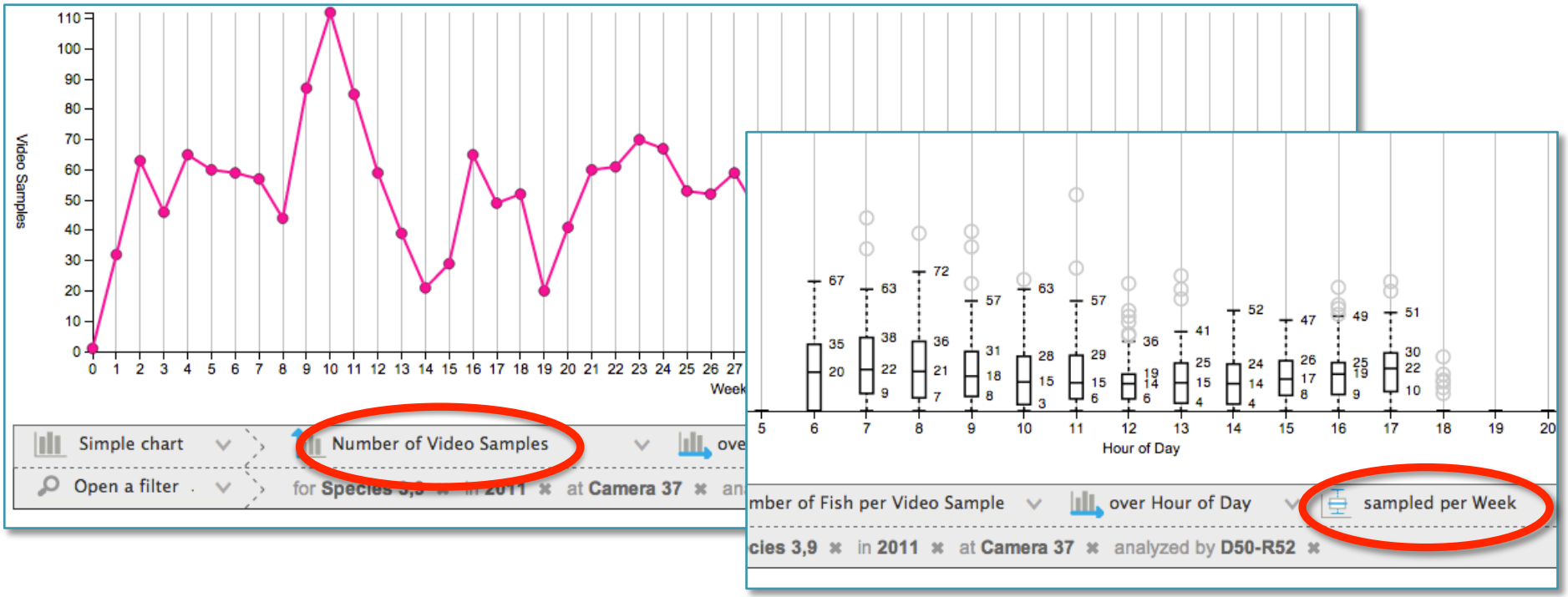


Access ground-truth images
Evaluation of biases, and rates of fish re-detection

Normalized fish count, with bias correction

- USER INTERFACE -

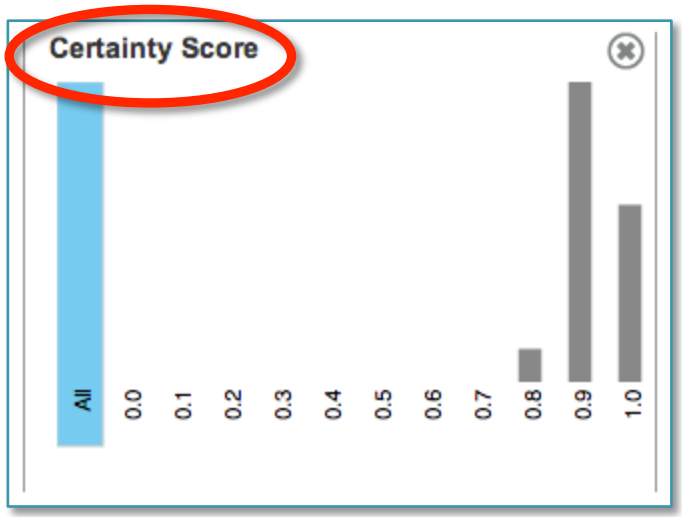
SUPPORT FOR UNCERTAINTY-RELATED INVESTIGATIONS



The 'Visualization' tab also addresses uncertainty issues

- **Number of videos** per location, time periods, image quality
- **Mean and deviation** of fish abundance and number of species

- USER INTERFACE -
SUPPORT FOR UNCERTAINTY-RELATED INVESTIGATIONS



The 'Visualization' tab also addresses uncertainty issues

- Fish abundance and numbers of species over **image quality**
- 'Certainty Scores' as an indicator of **video analysis accuracy**