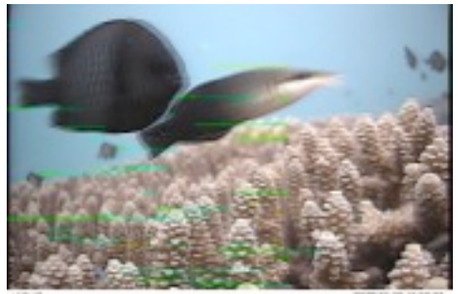
Experts, non-experts and automatic methods in crowdsourcing for wildlife image annotation

Jiyin He, CWI
soHuman 2012

The big picture

The Fish4Knowledge Project



Videos

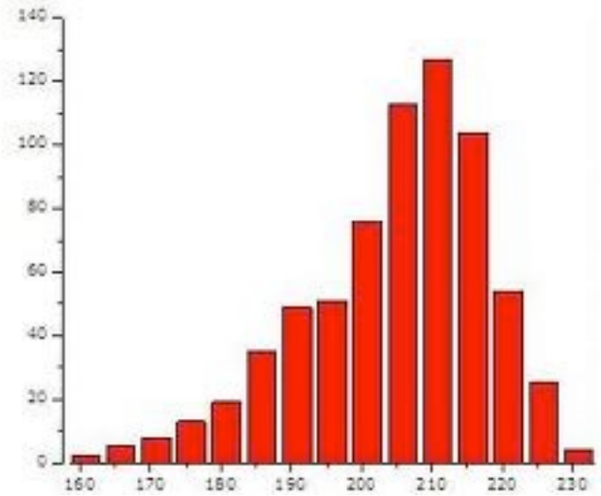
Underwater cameras

Computer vision systems

Detection

Tracking

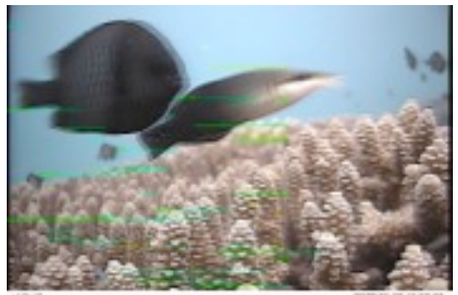
Recognition



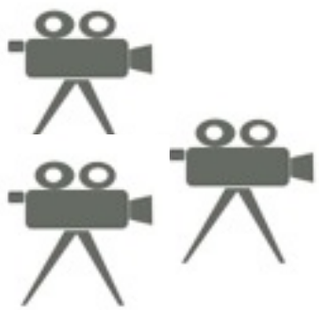
CWI

The big picture

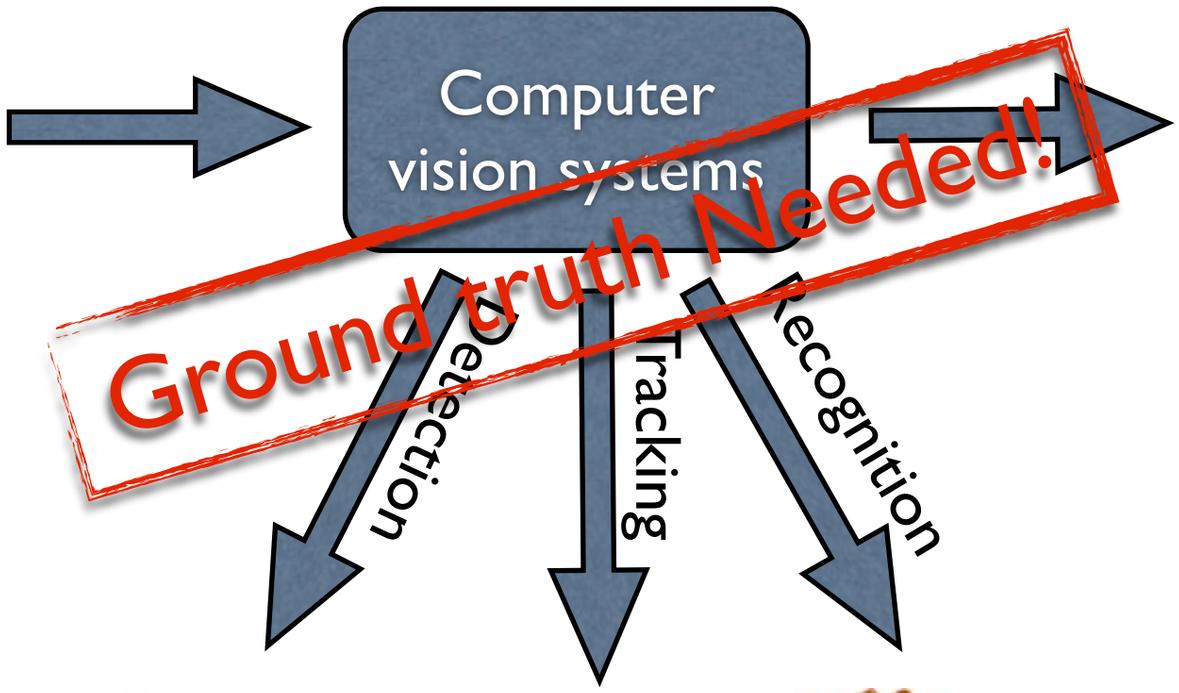
The Fish4Knowledge Project

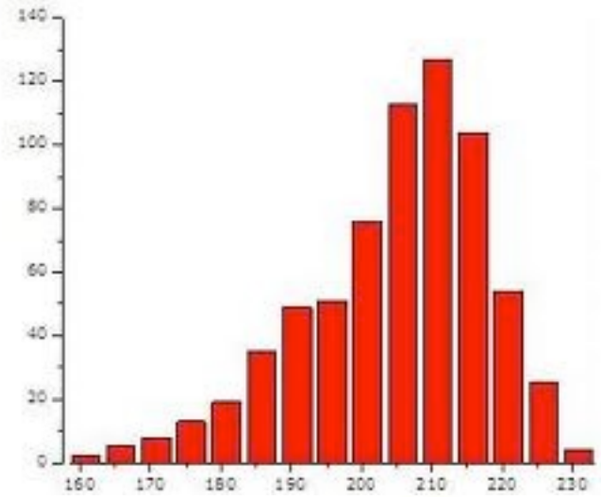


Videos



Underwater cameras





CWI

Fish recognition: a difficult problem

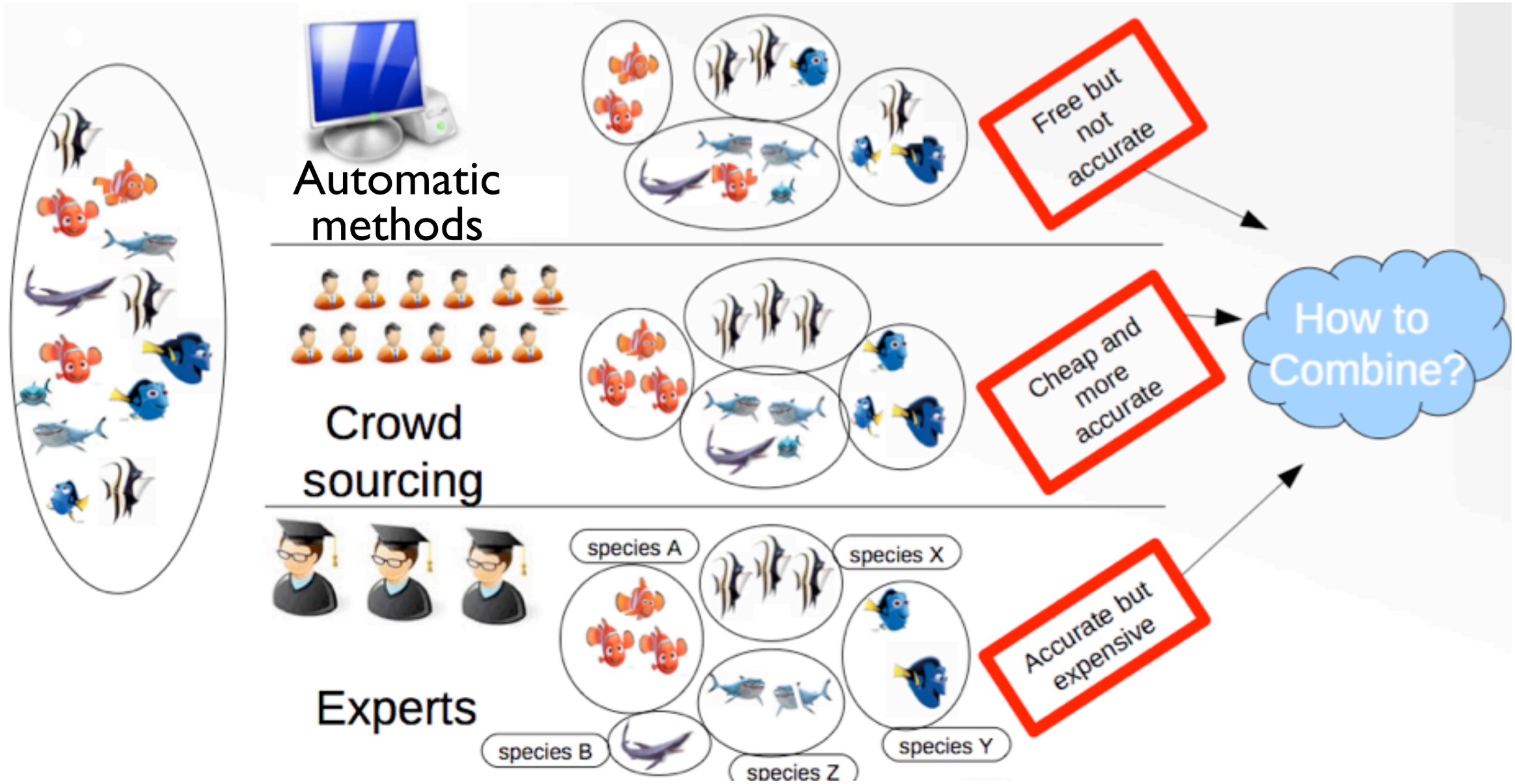
- Goal: given an image containing a fish, assign the species name of the fish to the image
- Large amount of images
- Expert knowledge needed
  - Non-experts often lack the knowledge needed to recognize a fish
  - Non-experts may not be able to map the common name of a fish to its scientific name
  - Even experts can have their expertise in different types of fish or fish in different areas
- Experts are expensive, rare resources

CWI

What can non-experts (not) do?

- Assumptions
  - Non-experts are not able to name fish species
  - But may be able to tell if two fish are visually similar
- Possible tasks
  - Manual clustering
  - Classification with textbook images as category labels

Automatic methods, crowds, and experts

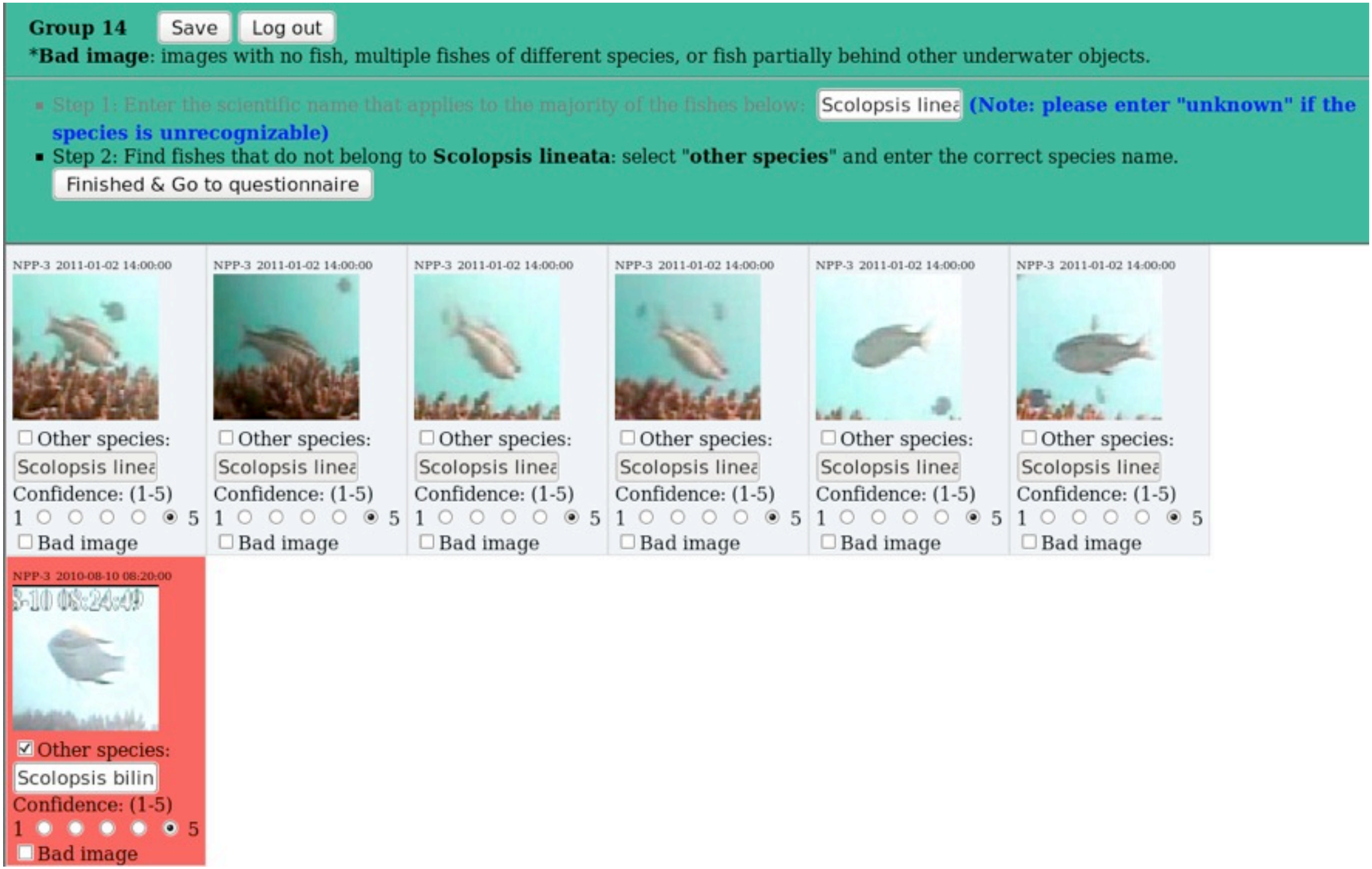


CWI

Using non-experts' effort to support expert annotation

- Cluster-based labeling
  - clusters constructed by non-experts
  - instead of labeling every image, experts label the clusters
  - correct the labels of individual images that are wrongly placed in a cluster
  - worst case: labeling every image

Interface (1)


Group 14
Save
Log out
*Bad image: images with no fish, multiple fishes of different species, or fish partially behind other underwater objects.
Step 1: Enter the scientific name that applies to the majority of the fishes below: Scolopsis linea (Note: please enter "unknown" if the species is unrecognizable)
Step 2: Find fishes that do not belong to Scolopsis lineata: select "other species" and enter the correct species name.
Finished & Go to questionnaire
NPP-3 2011-01-02 14:00:00
Other species:
Scolopsis linea
Confidence: (1-5)
1 5
Bad image
NPP-3 2010-08-10 08:20:00
Other species:
Scolopsis bilin
Confidence: (1-5)
1 5
Bad image

CWI

Interface (2)

- A questionnaire after labeling each cluster
  - How difficult was it to recognize the fish (1-5 scale)
  - What makes the recognition difficult?
  - What helps recognition?
  - Are the “other species” related to the main species in the taxonomy tree?
  - What are the most discriminative features of the main species in the cluster?

Expert annotated data

- 3 marine biologists with over 10-25 years' research experience in Taiwan sea area.
- 27 manually constructed clusters
  - For each cluster, at most 30 images are randomly sampled to be shown to the biologists

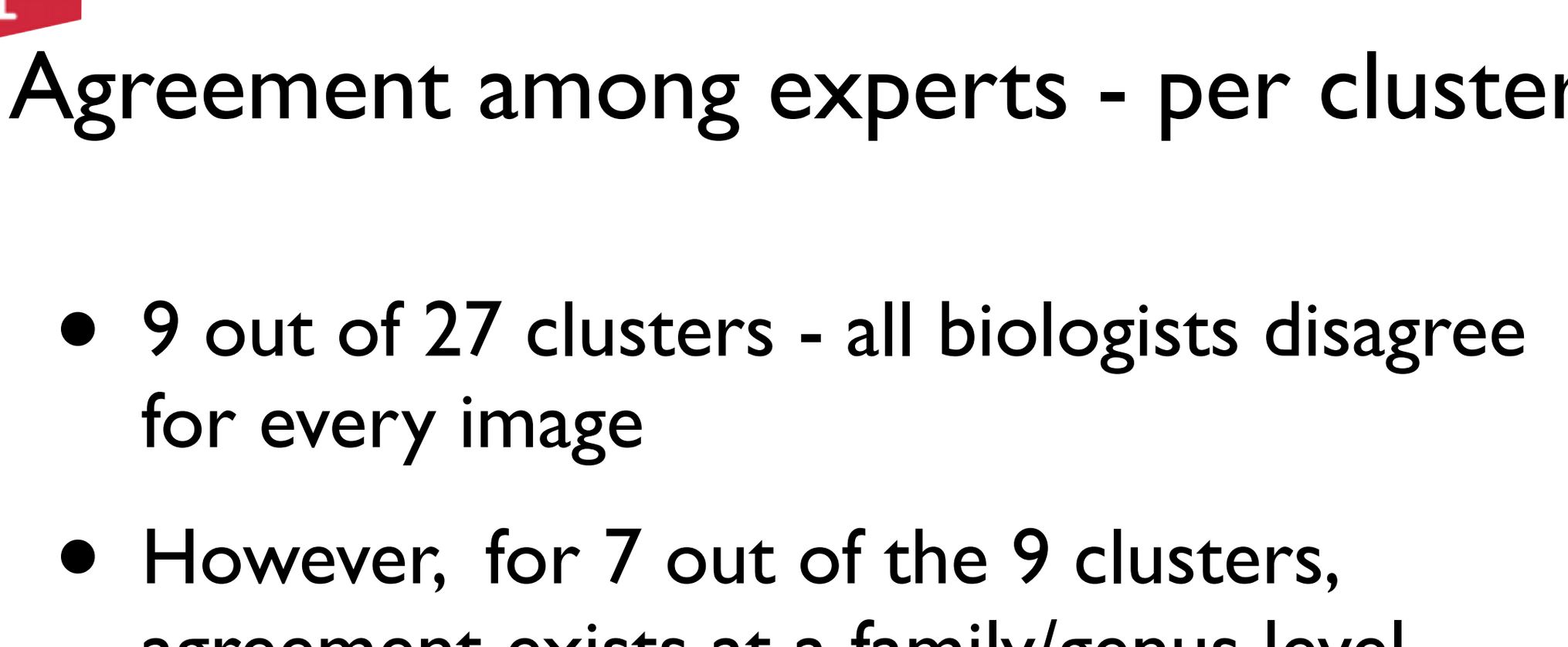
CWI

Agreement among experts - image level

- 82.6% images - at least 2 biologists agree on a name
- 56.3% images - 3 biologists agree on a name
- Note: sizes of clusters are not evenly distributed

Agreement among experts - per cluster

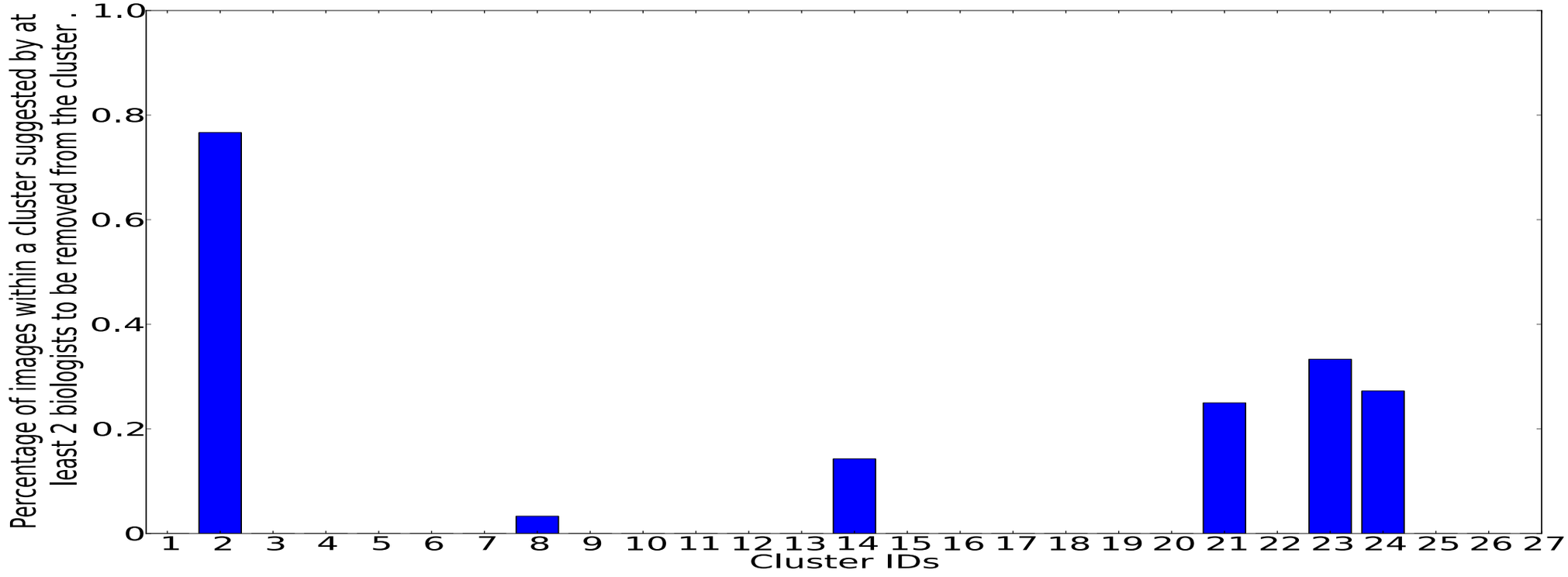
- 9 out of 27 clusters - all biologists disagree for every image
- However, for 7 out of the 9 clusters, agreement exists at a family/genus level



||

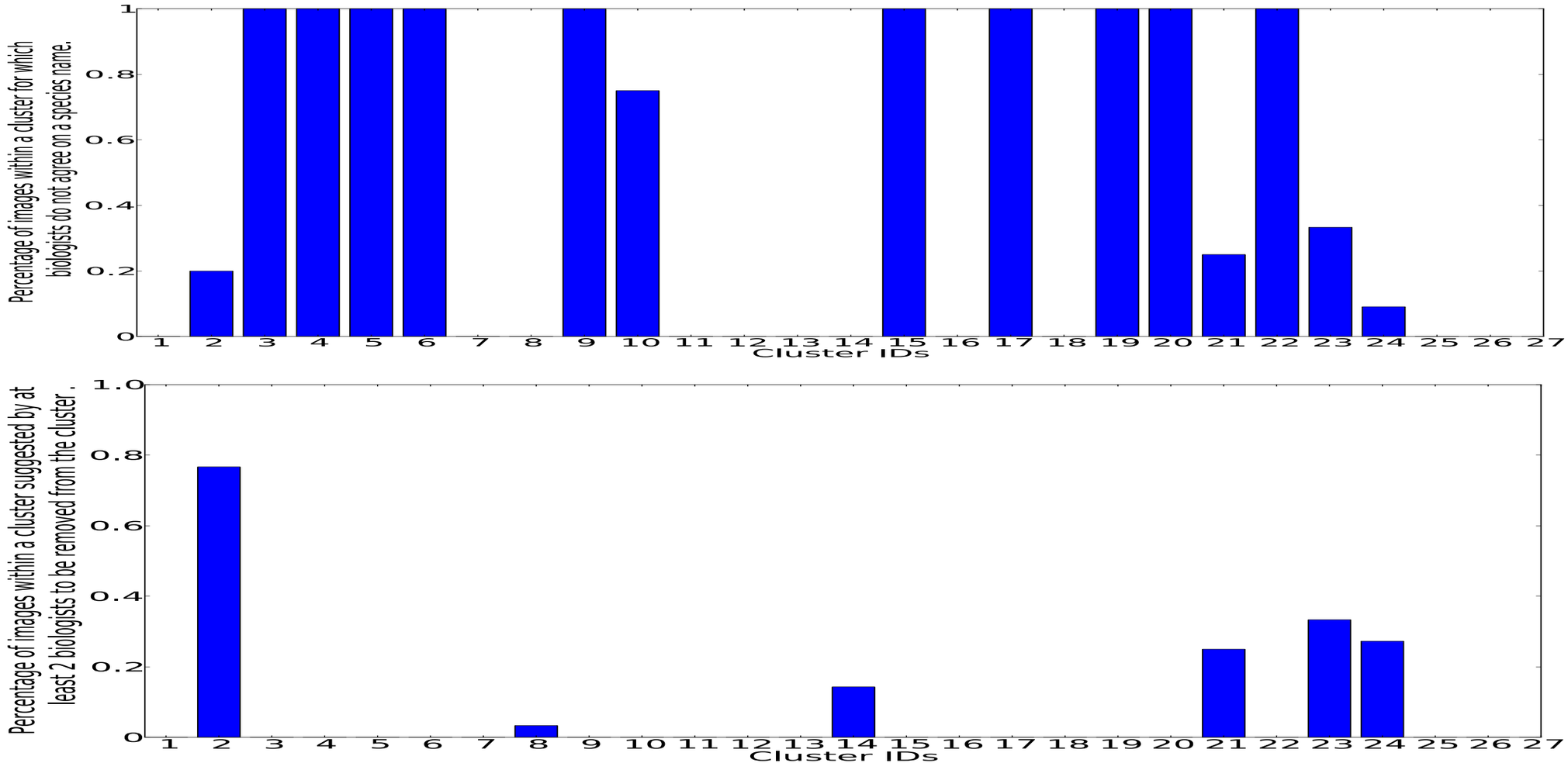
Performance of non-expert manual clustering

- 6 out of 27 clusters contain "other species"



Experts vs. non-experts

- Difficult for experts ! = difficult for non-expert
- Clustering != Recognition



Questionnaire

- What makes recognition difficult?
  - 21/27 cases: low resolution
  - 17/27 cases: there exist very similar species

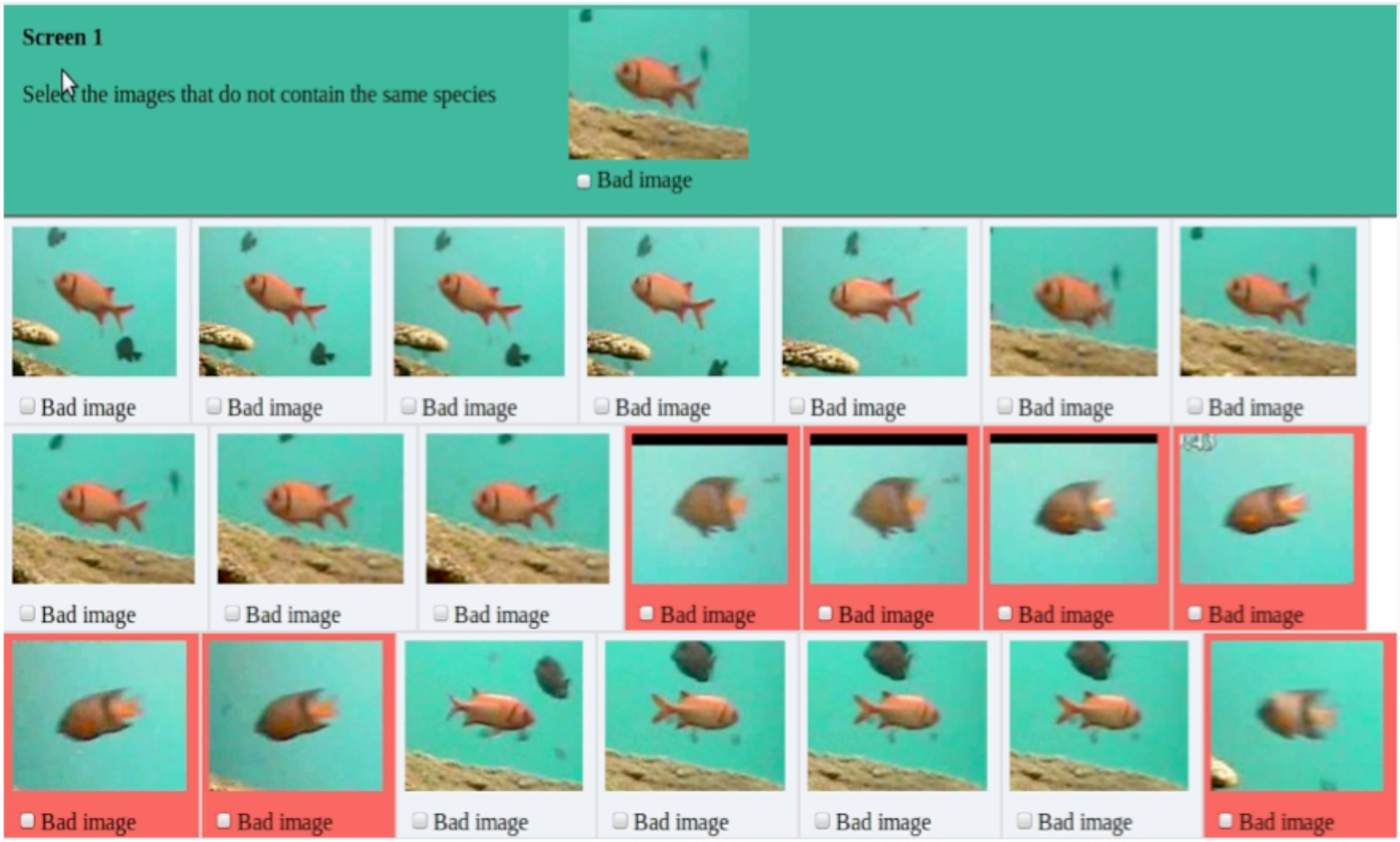
- What helps recognition?
  - 24/27 cases: features of the fish
  - 17/27 cases: experience
  - 5/27 cases: location
  - 3/27 cases: better resolution

CWI

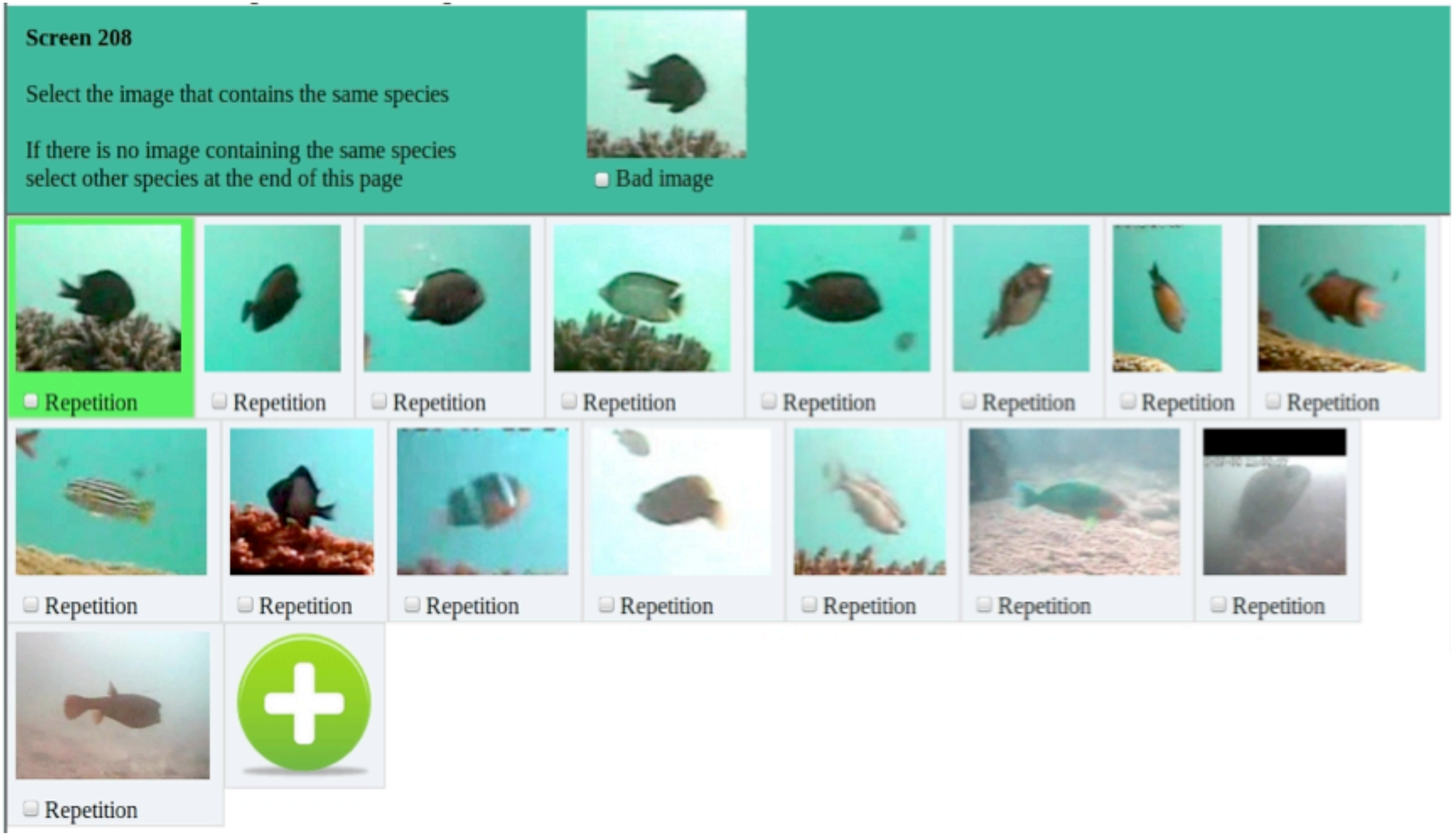
Lessons learnt

- Cluster-based labeling approach enables experts to label a relatively large amount of images with a limited amount of effort
- Non-experts are able to measure the visual similarity between fish images, thus clustering or classification with visual labels are possible
- Gap exists between clustering and recognition - nontrivial even for experts

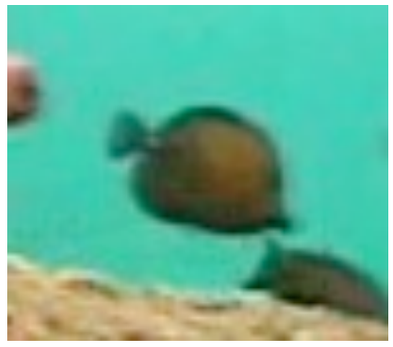
Using automatic method to support non-expert annotation: cluster-validation (1)

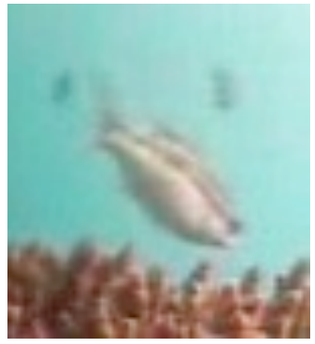


Using automatic method to support non-expert annotation: cluster-validation (2)



How many different species are there?





How many different species are there?

Dascyllus reticulatus

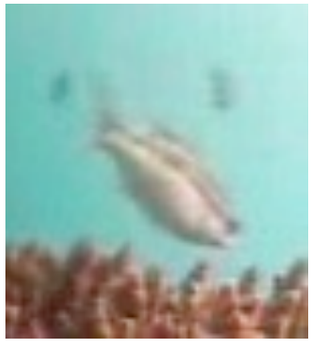
Acanthurus sp.

Zebrasoma scopas

Acanthurus sp.

Scolopsis bilineata

Scolopsis lineata



Scolopsis lineata

Further steps

- Bridging gap between clustering and recognition
  - Linking clusters to species names
    - Using crowd votes to refine the annotations of images that are visually similar
- Online learning to combine the effort of automatic methods, crowd and experts