Detecting, tracking, and classifying visual events in underwater video:

Automatic Visual Event Detection (AVED)

Duane Edgington

Problem we are trying to solve

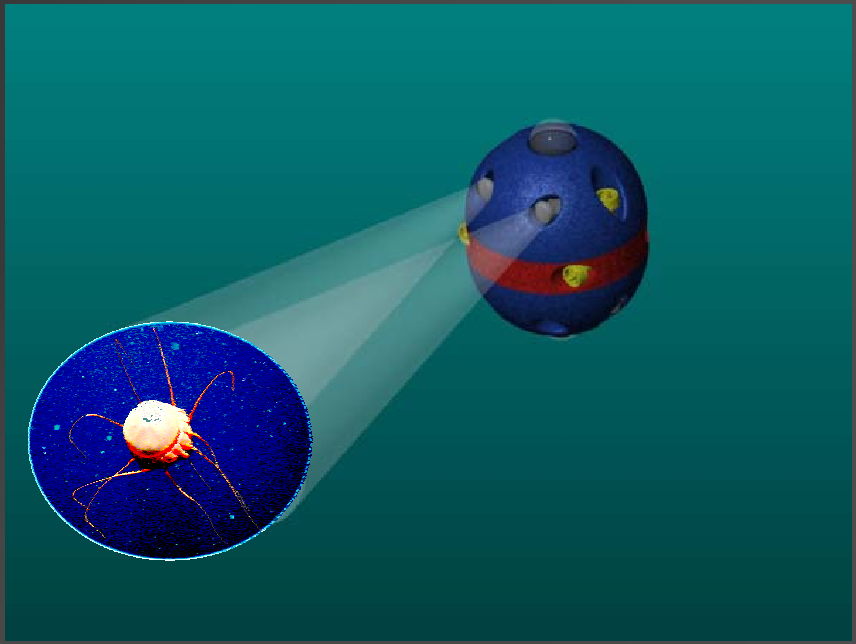
- MBARI ROVs have proven that high quality video is a useful quantitative scientific instrument for oceanography
- BUT today analyzing and interpreting science video is a labor intensive human activity
- Scaling from 2 ROVs to multiple AUVs to 100s of Observatory cameras cries out for automated video analysis

ROV Video processing

- MBARI Collects ~300 days per year of video from its ROVs
- 13,000+ tapes & 9,000+ hours of undersea video
- Need to enable integration of results over many dives over many years.
- Annotating video is time-consuming and tedious. Can we supply tools to make the analysts more productive and their lives easier?
- Can we build systems for real time analysis & event response at sea?

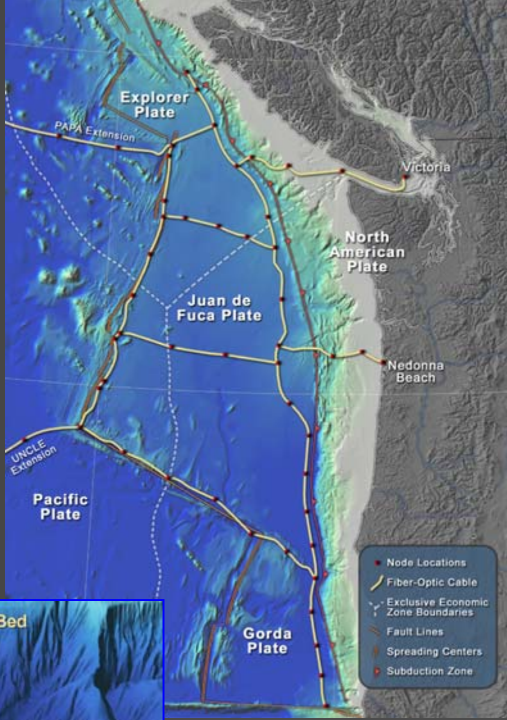
Quantitative video / tracking on AUV

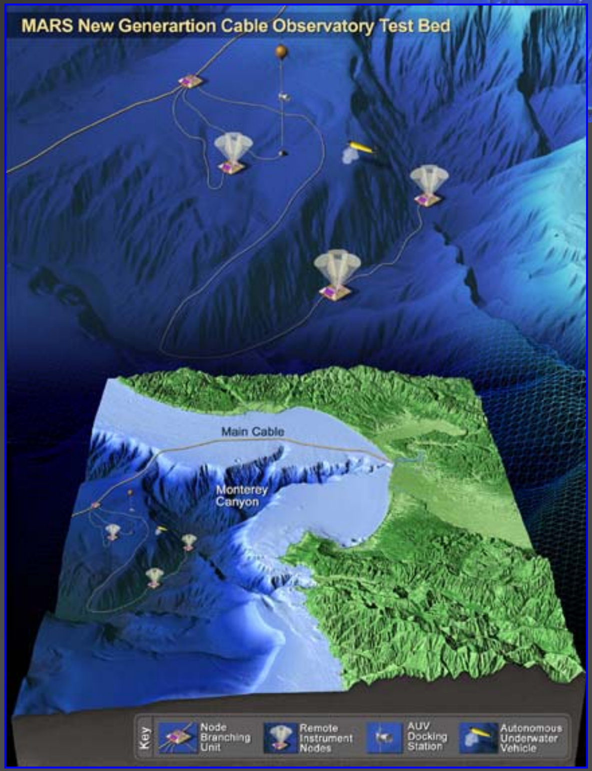
- 2003: technology explorations looking at video cameras on AUVs
  - Quantitative Low Light Imaging Module: use cases for midwater video transects
  - Deep Ocean Tracking Instrument: use cases for active underwater surveillance



Cabled observatories

- Use cases to install fixed video cameras in underwater observatories to record data 24/7 – Prototype: Monterey Accelerated Research System (MARS)
- Each camera would require 4-5 full time video analysts to manually capture information in tapes.
- How will we deal with 10's or 100's of live video feeds from cabled observatories?
- Event response (pan/tilt/zoom to events





Midwater video transect

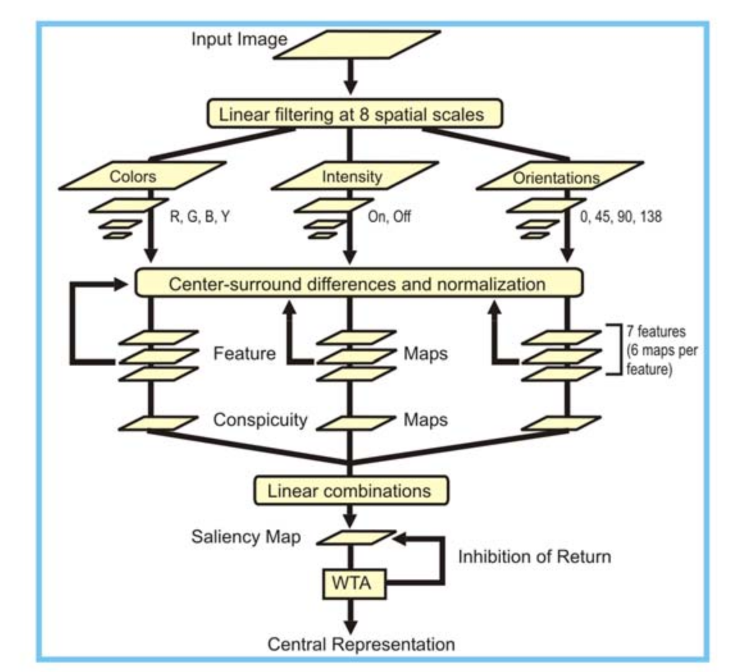
- (external viewer)

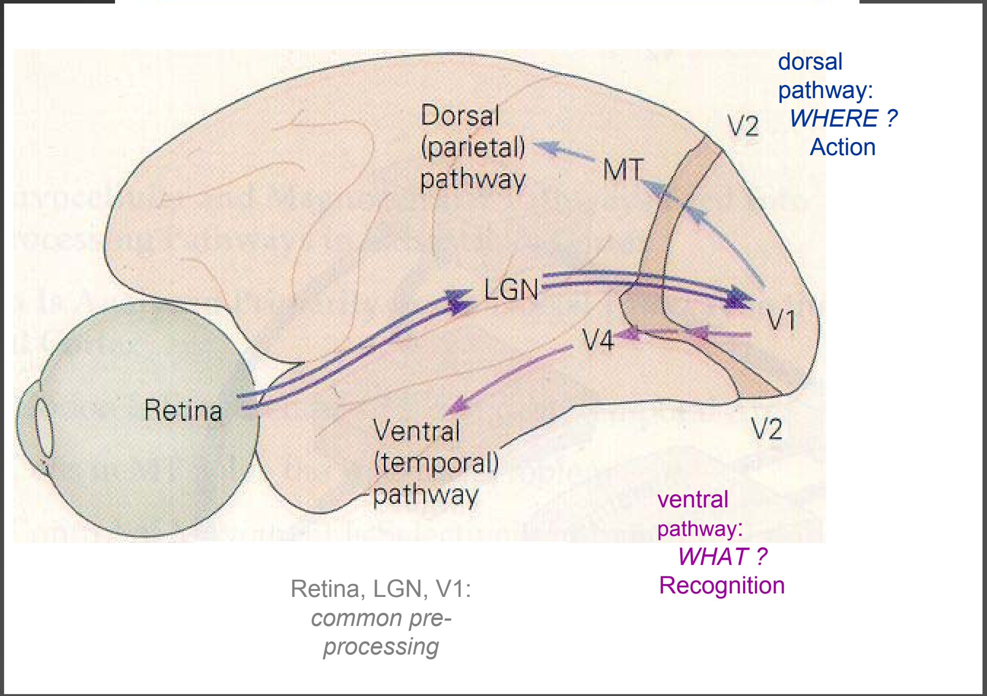
Approach to the problem

- We know that humans can detect and identify animals in the underwater video. How do humans do it? Clues from neurobiology?
- Starting in 1999: survey machine vision and artificial intelligence algorithms for natural scene video image analysis
- 2001-2002: zeroed in on "Neuromorphic Engineering" approach to artificial vision
- 2002-now: partnered with research labs (Caltech and USC: Koch, Perona and Itti).

Core approach

- Saliency model of attention – detecting events in the visual scene (Koch and Itti)
- Early Vision model of recognition – analysis of image sequences to recognize objects (Perona)





General outline of processing

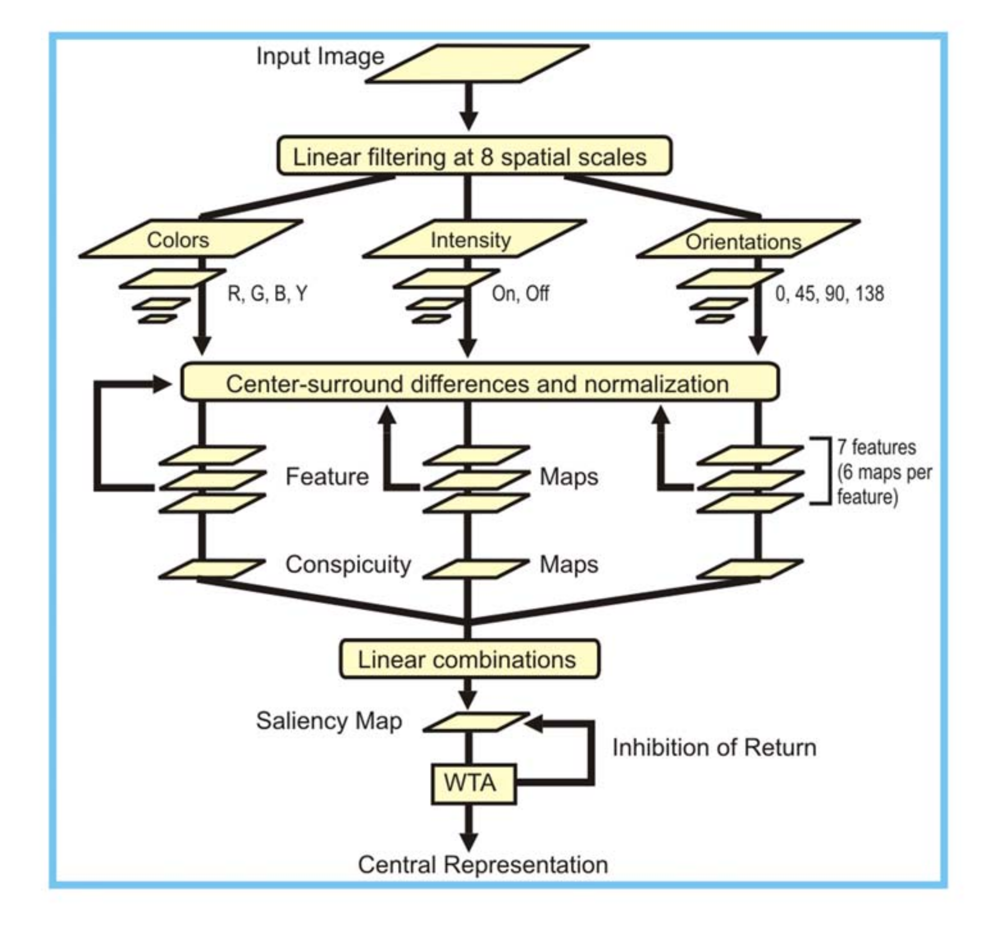
- Preprocessing
- Detection
- Tracking
- Classification
- Visualization

Can you spot the siphonophores?

Rob could spot the siphonophores…

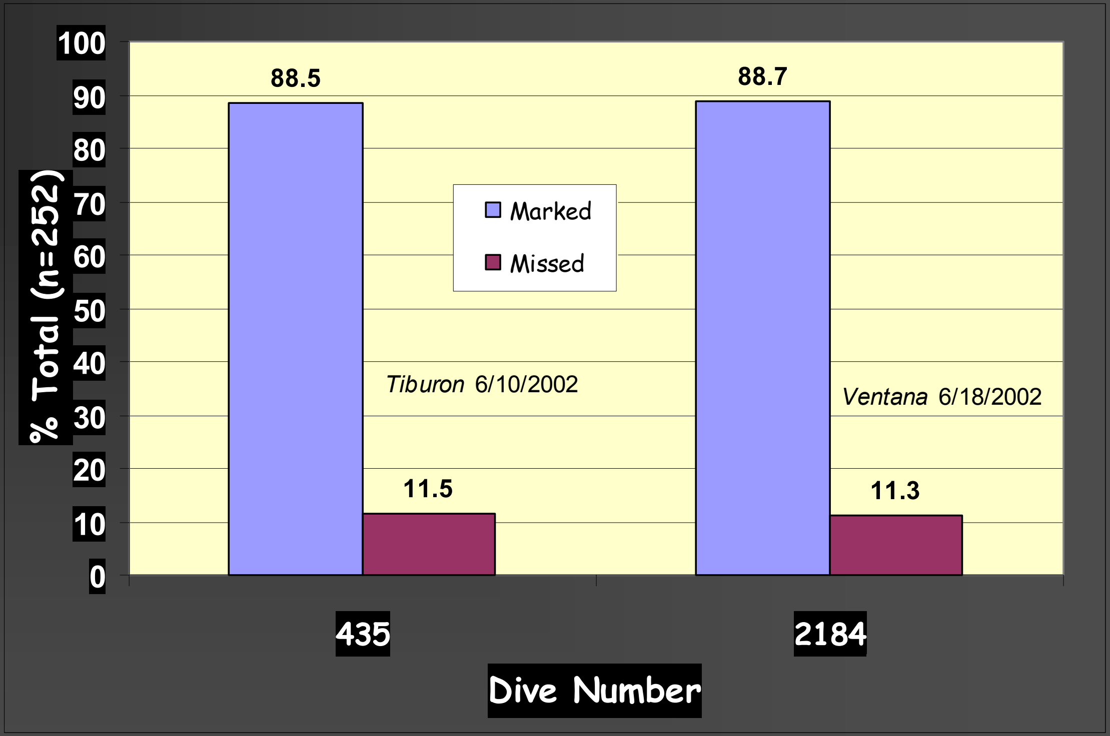
Detection

- Look for strong signals in the scene:
  - Color contrast
  - Illumination contrast
  - Edges
- Select strongest of these signals
- Midwater: bias for edges, color
- We call the objects yielding the strongest signals "salient"



Does the algorithm detect animals in the scene?

- Analyzed several hundred video frames from midwater transects:
  - Animal in the scene?
  - Yes: What was the most salient event marked by the algorithm?
    - Animal?
    - Snow?

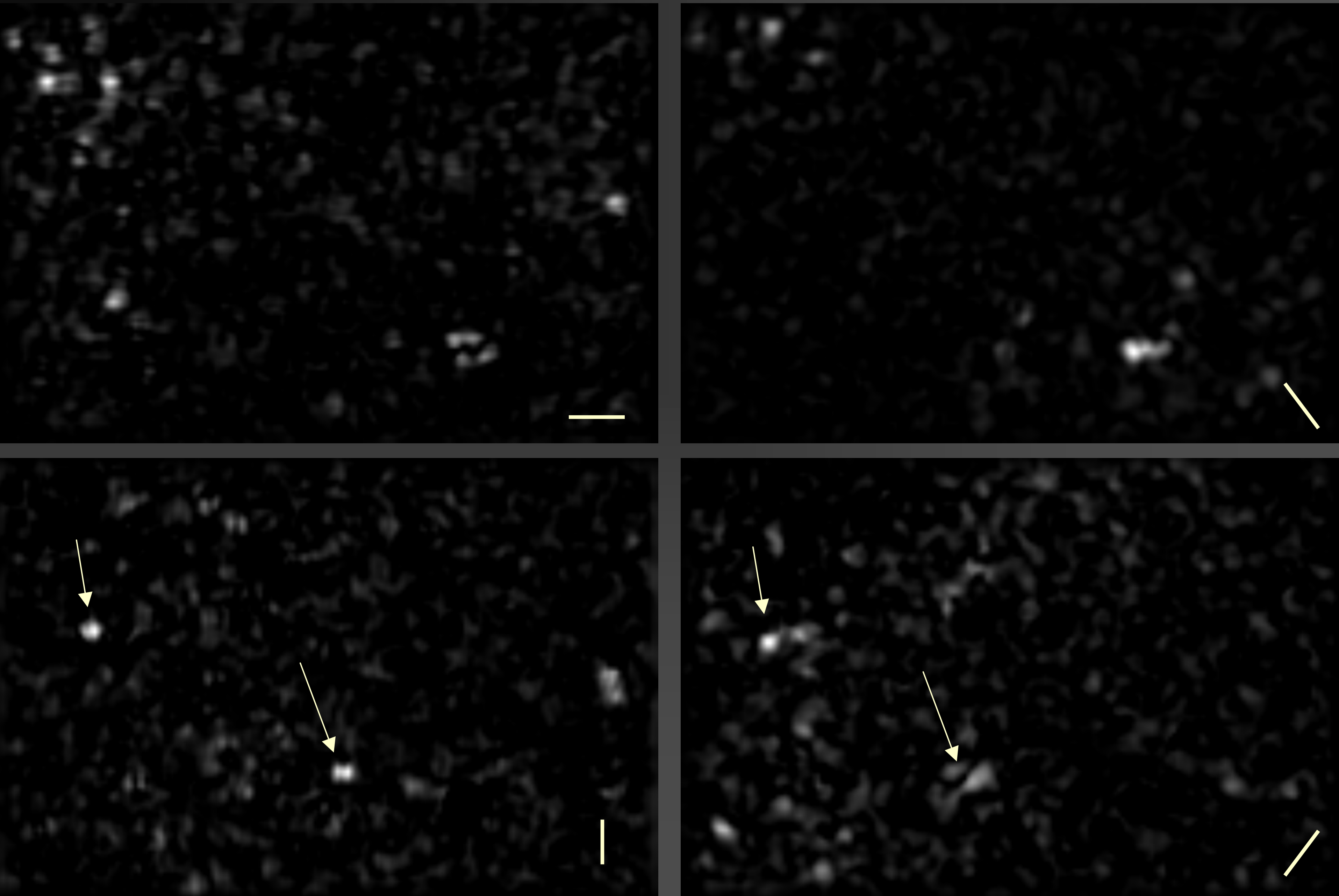


Detection – combine with tracking

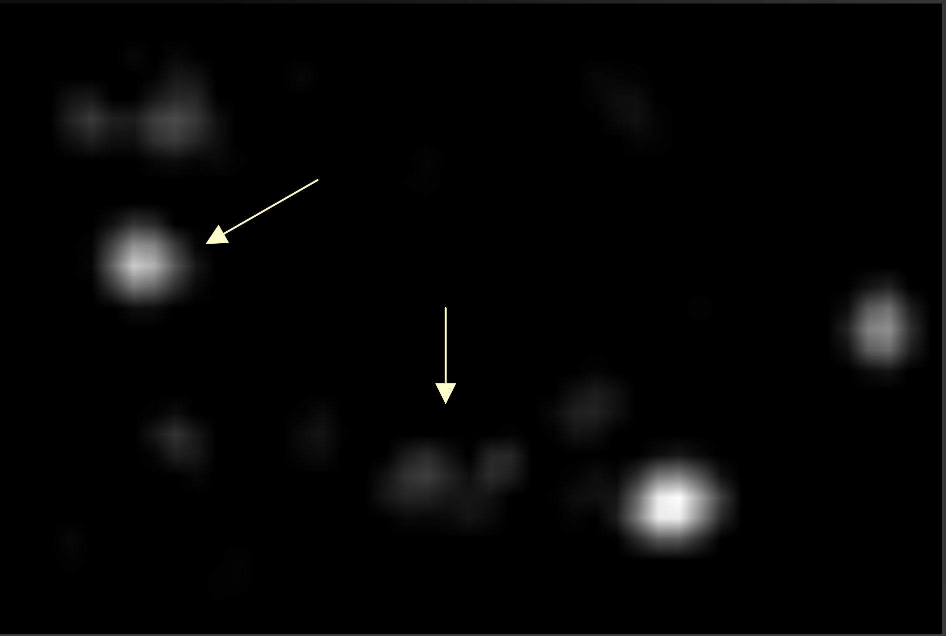
- Use the saliency-based attention model
  - Track only salient objects
    - Keep # of tracked objects relatively small
  - Classify as "event" based on persistence and degree of interest
- Problem:
  - Low-contrast elongated structures (siphonophores) are often eclipsed by high-contrast "snow"
    - Enhanced saliency-based attention model with a model of lateral inhibition

Our scene again…

Orientation filters

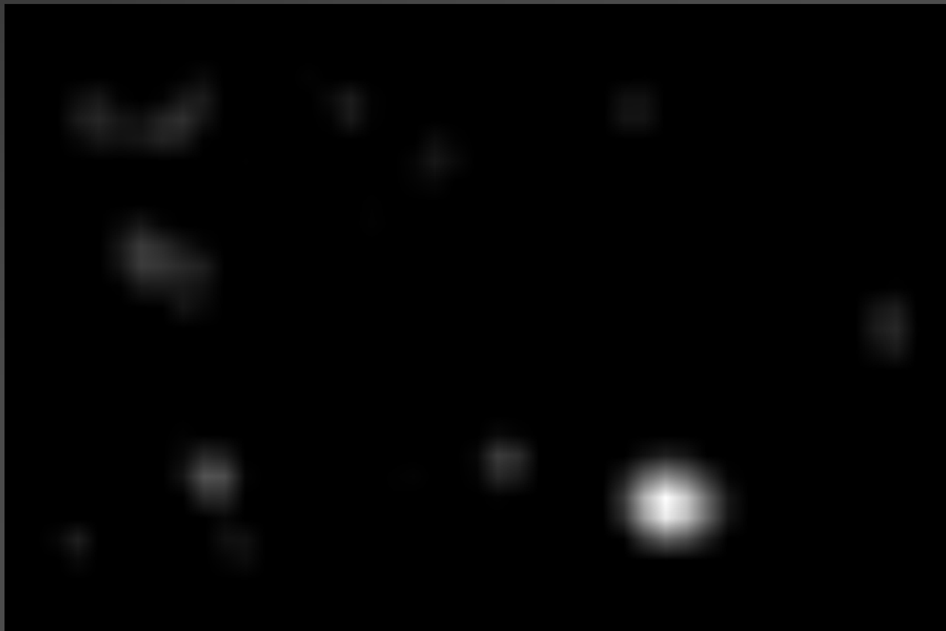


Feature map with inhibitory interneurons



Stronger signal for the two faint siphonophores with lateral inhibition

For comparison: feature map without inhibitory interneurons:



Tracking using Kalman filters

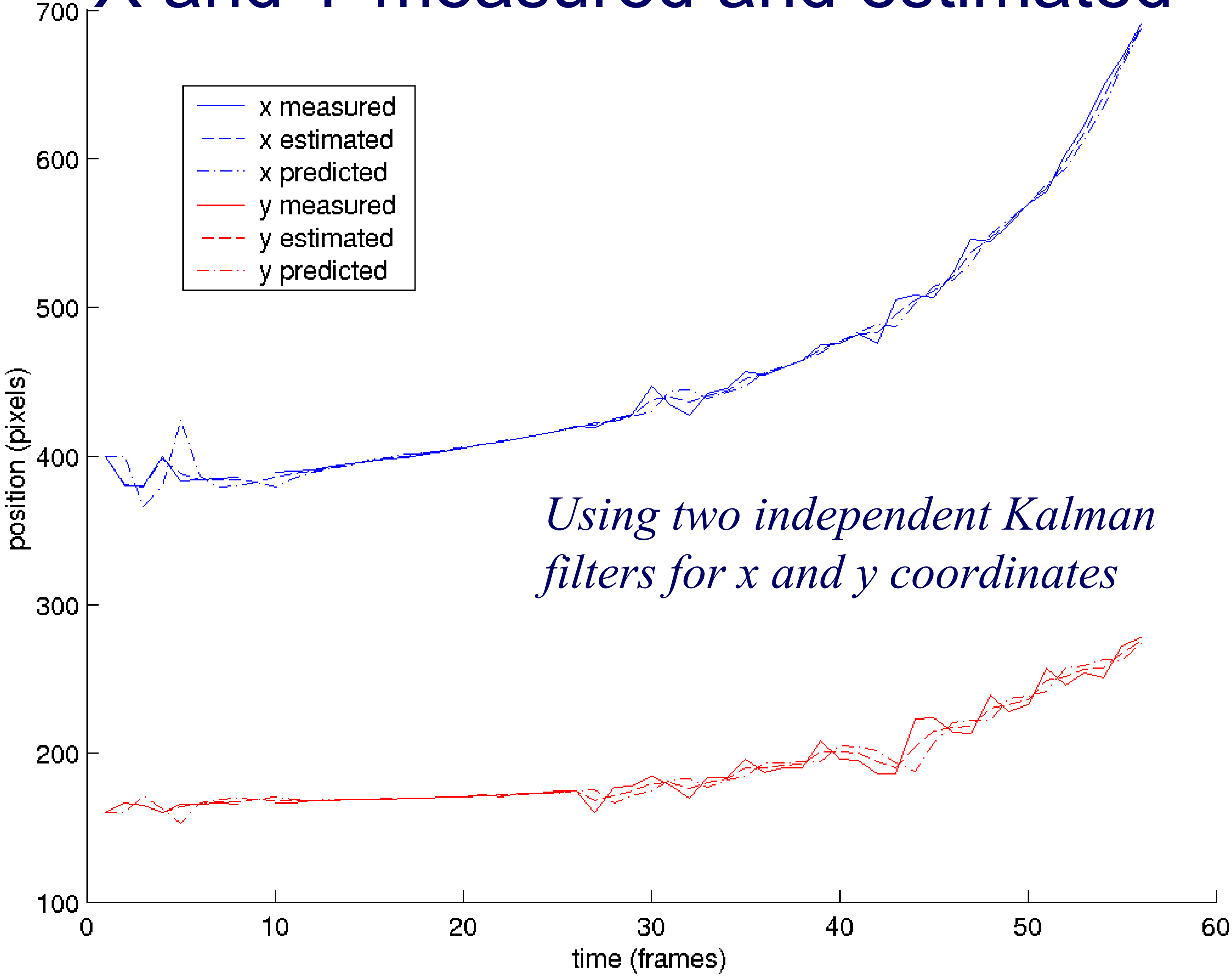
- Track based on prediction of trajectory (optical flow)
- Assign salient objects to trackers
  - Manageable since we don't track too many objects at once

Kalman filters

- A way to predict future positions of an object in the scene
- Special trick of Kalman filters: Given:
  - The uncertainty of how well the model describes reality, and
  - The uncertainty of the measurements,

  predict the next location
- Kalman Filter also enables extrapolation across missing measurements (fuzzy image in a frame)

X and Y measured and estimated

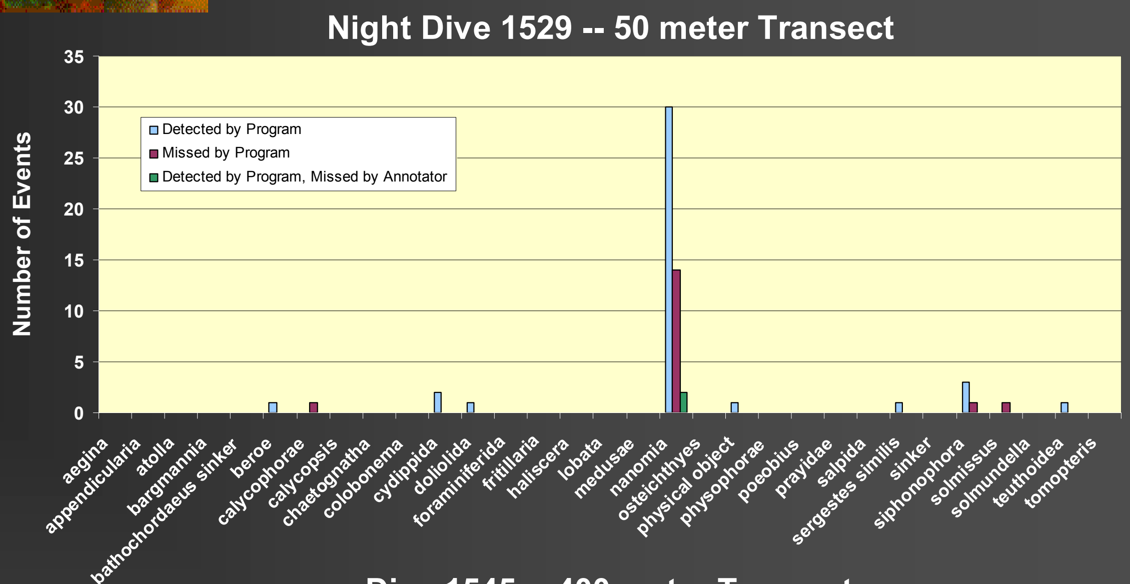


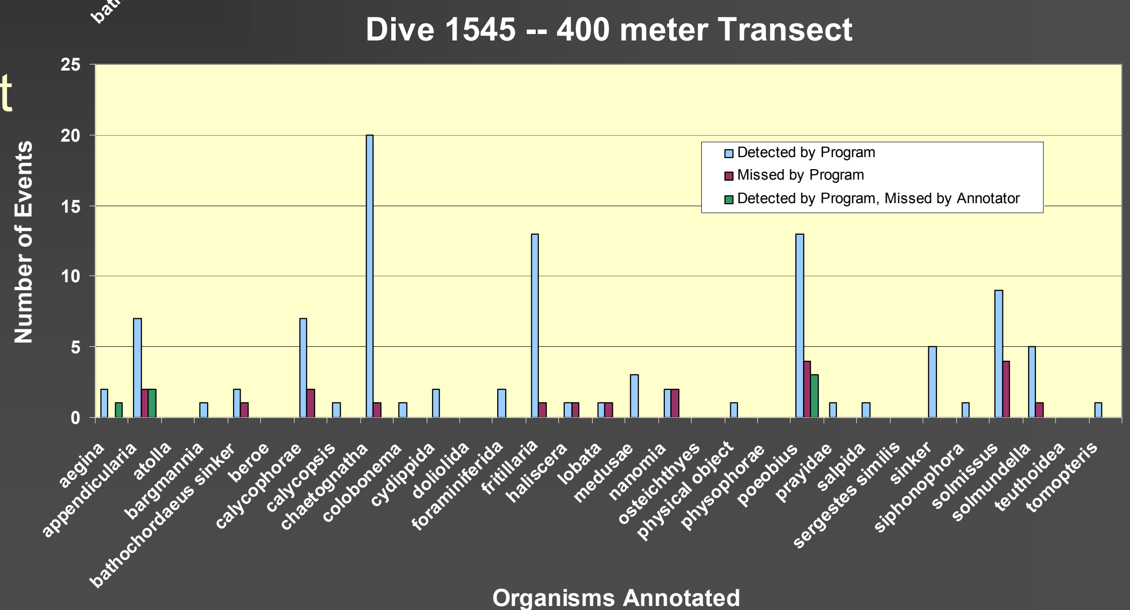
Using two independent Kalman filters for x and y coordinates

The annotated movie

Comparing detected events with annotation

- Analyzed midwater video transects that had been fully annotated
- Asked:
  - What percentage of annotations did the program detect?
  - What percentage did it miss?
  - Did the program detect any animals that the annotator missed?
  - What else did the program detect?

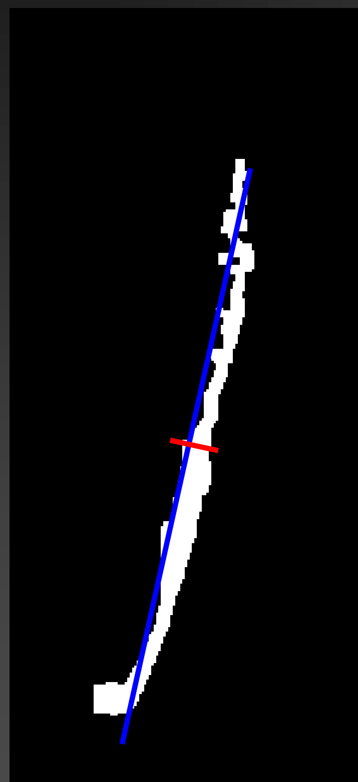




Classification

- Explored classification into “interesting” and “boring” according to examples provided by professional video annotators at MBARI

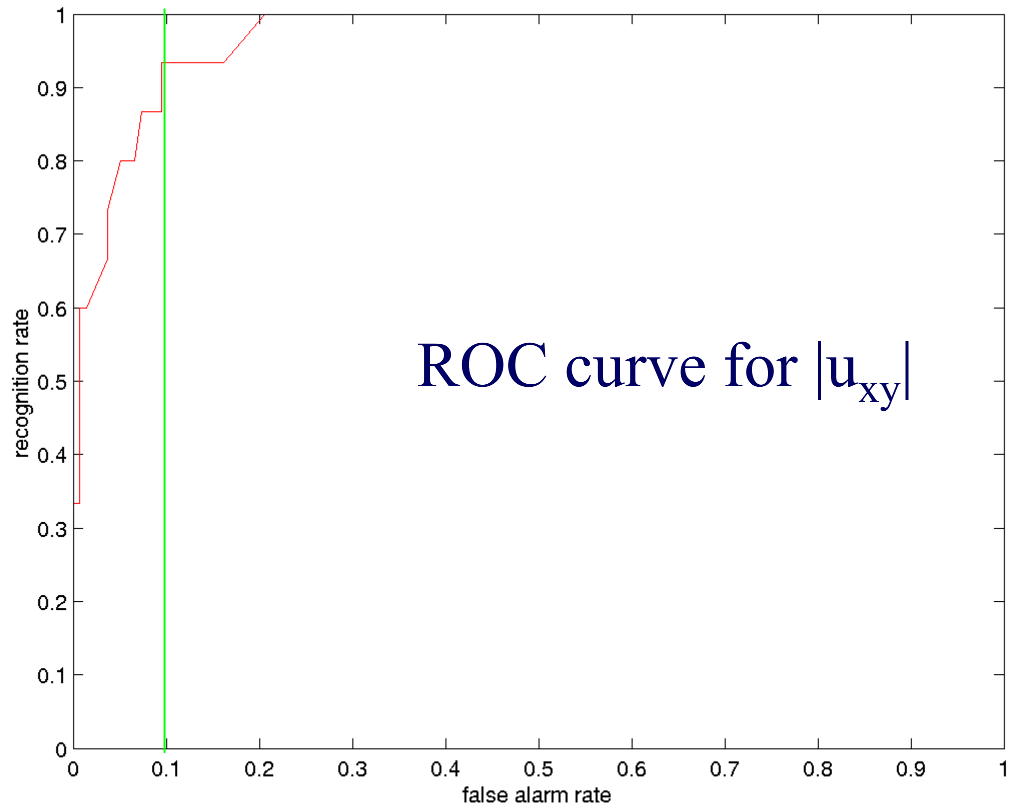
Measuring object shape properties



- 11 Properties extracted:
  - Area (in pixels)
  - Second moments with respect to the centroid u_{xx}, u_{yy}, and u_{xy}
  - Major and minor axes
  - Elongation and orientation
  - Mean, minimum, and maximum intensity over the object shape

Classification

- In a preliminary attempt, only thresholding $|u_{xy}|$ yielded 93% correct classifications with 10% false positives

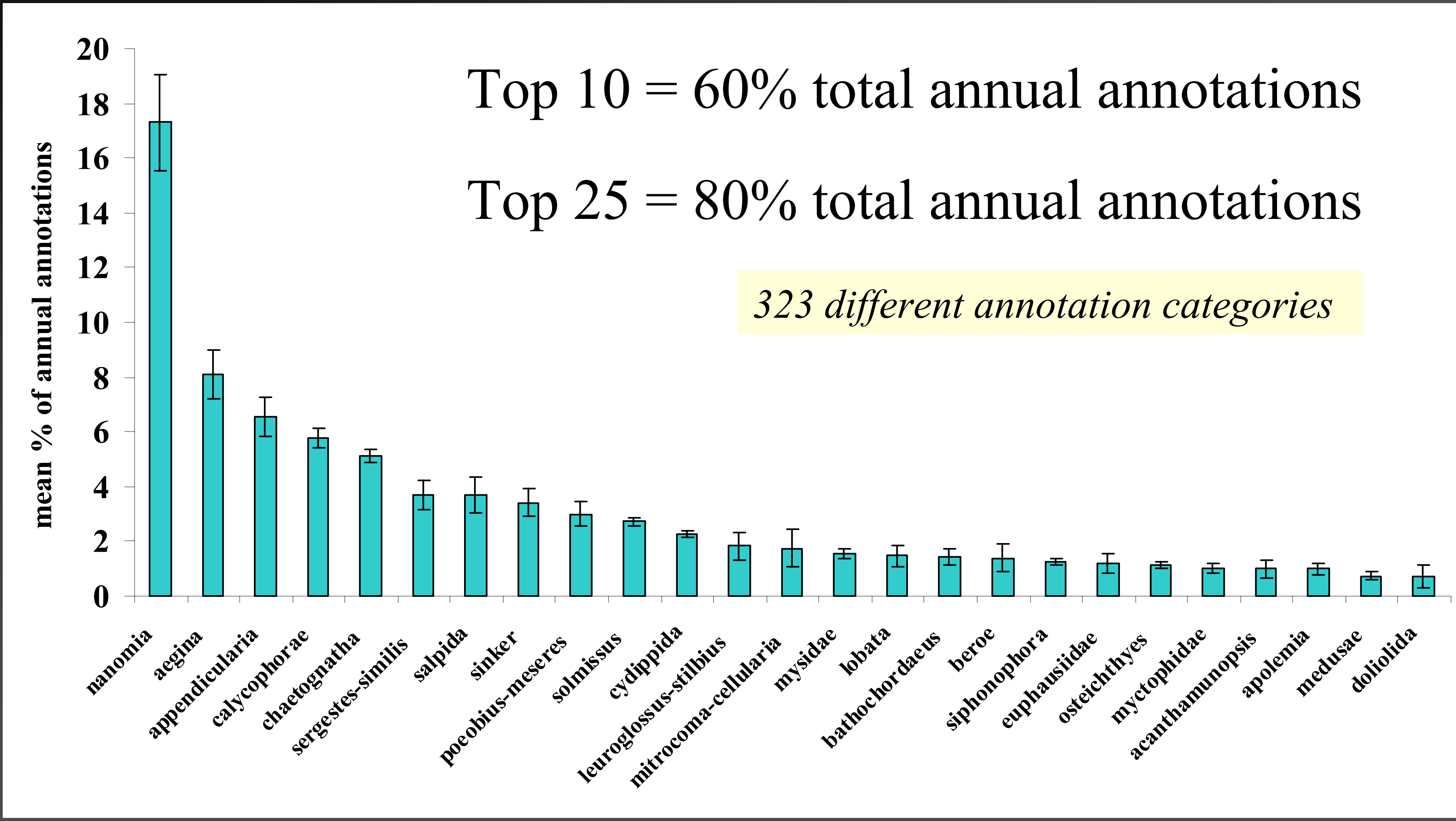


- After discussion with video analysts: classification into “interesting” and “boring” is not useful – system only makes since if it recognized species or at least families of animals reliably

Classification – statistics of animals

- With 10,000s of animals this seems daunting. Is it?
- 2003 MBARI Summer Intern Alexis Wilson (Middlebury College, VT) mined the Video Information Management System (VIMS) database from 1997- 2002
- Collected data on ALL organisms in top 1000m of the water column
- Ranked organisms according to percentage of annual annotation frequency

Results: top 25 organisms



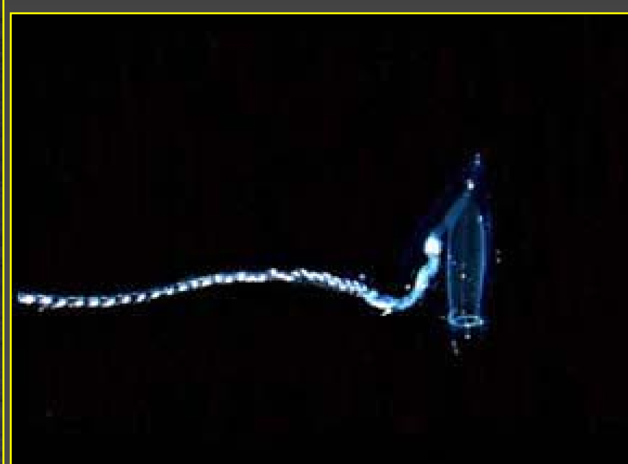
Major Players

#1 Nanomia (17.3%)

#2: Aegina (8.1%)

#3 Appendicularia (6.5%)

#4 Calycophorae (5.8%)



#5 Chaetognatha (5.1%)

Classification

- Collected example video clips for each of the 25 most common animals
- An approach: step by step develop classifiers for the most common animals
- If we manage to classify the most common ones, we'll able to automate some of the annotations
- Being able to recognize specific kinds of animals will help with research programs such as abundance counts

Motion as a feature

- Motion can be a very important feature in at least two ways:
  - Objects that do not follow the general optical flow caused by the ROV motion apparently propel themselves – good indicator for an animal; this is work in progress
  - The way an animal changes its shape while it is moving is very distinctive for certain species – annotators use this a lot for classification; this problem is on our future "to-do" list

Visualization

- Currently: mark the objects with colors in the video frames – either with a bounding box or by their outline
- Eventually, automatic fast-forward to the next salient event would be possible
- Extraction of single objects for closer study

Example: marking the outlines

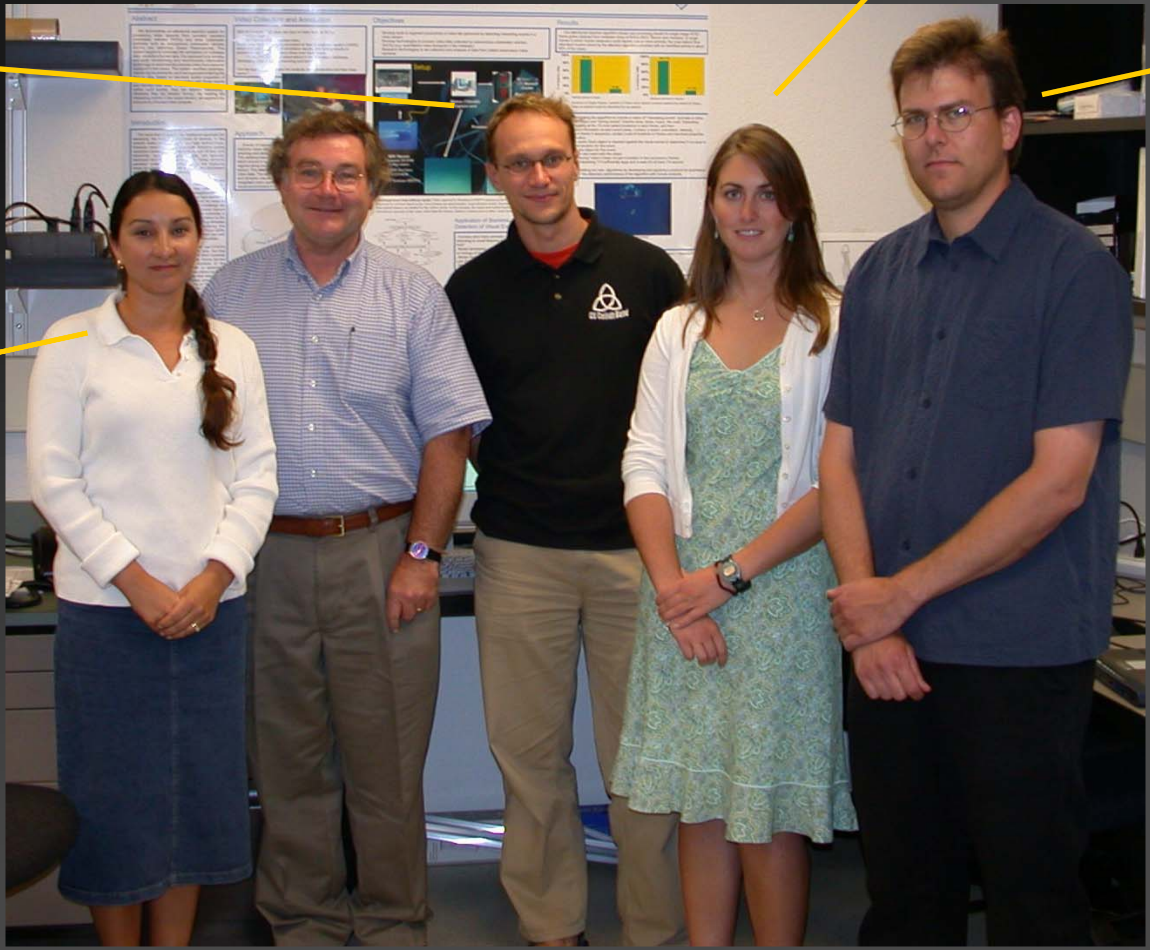
Contributors include…

Alexis Wilson

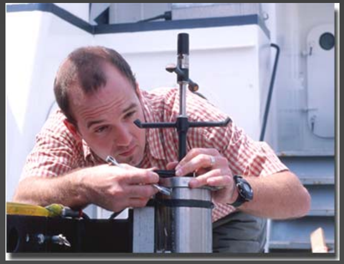
Dirk Walther (Caltech)

Mike Risi

Danelle Cline



Rob Sherlock



Karen Salamy