The Case for 3D Imaging

Hank Bart

Tulane University Biodiversity Research Institute

hbartjr@Tulane.edu

Taxonomic Study

- Goal of taxonomic study of specimens is to identify characters that uniquely diagnose a group of specimens under study (suspected undescribed species) as distinct from other (known) groups.
- For ages, taxonomists have engaged in laborious and time consuming efforts to identify and quantify these characters.
- Advent of the information age gave taxonomists and other biologists hope that machines might expedite the data gathering and analysis processes.

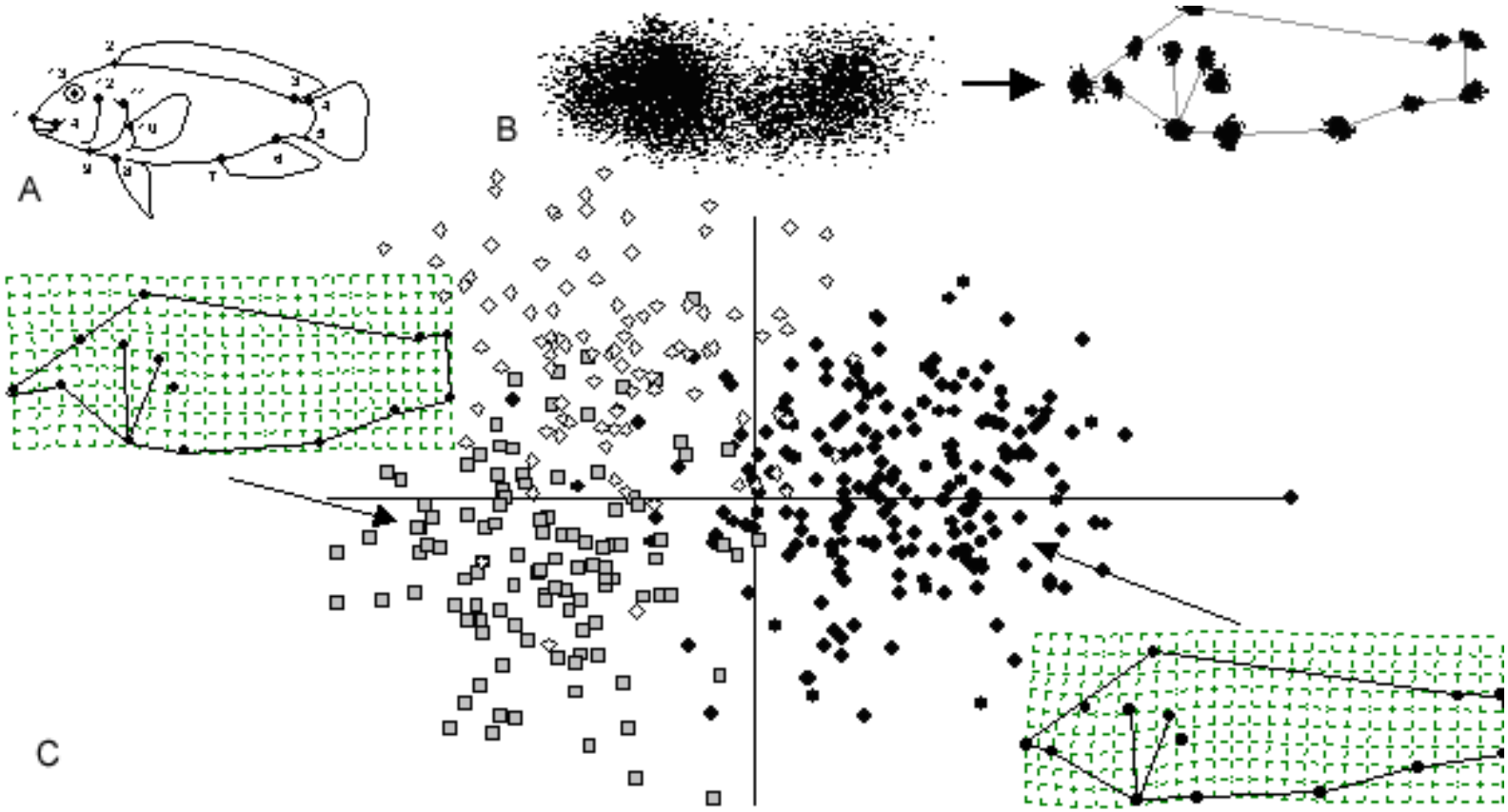
Traditional Methods

- Counts (meristics) and measurements (morphometrics) on specimens, pigmentation patterns, other features.
- Traditional approach to morphometric analysis usually involves application of multivariate statistical procedures to collections of distances, angles, or distance ratios.

Geometric Morphometrics

- Analysis of linear distances, ratios, or angles frequently fails to capture the complete spatial arrangement of the anatomical points on which the measurements are based.
- Assigning Cartesian coordinates to the points (landmarks) used to define traditional measurements concisely encodes all the information in any subset of distances or angles between the landmarks.

Geometric Morphometrics



http://life.bio.sunysb.edu/morph/review/review.html

GM Limitations

- GM preferred method of investigating shape change; tells taxonomists parts of the body to focus on to quantify differences.
- Variables derived from GM methods (partial warps) are not directly interpretable in terms of particular aspects of body morphology; quantification requires additional work and time on the part of taxonomists.

Machine Learning

- Subfield of artificial intelligence.
- Employs computer algorithms that improve their performance over time (“learn”) through training and testing process involving various forms of data.
- Closely related to fields of data mining, statistics, inductive reasoning, and pattern recognition.
- I have been collaborating with computer scientists since 2005, applying machine learning techniques to fish taxonomy.

Machine Learning in Taxonomic Research

In 2008, we proposed to build a cyberinfrastructure for machine aided fish taxonomy research with two components.

- Species Identification
  - Images of categorized specimens representing *known* taxonomic groupings of fishes used to train a collection of statistical models to identify fish (family and species levels).
- New species discovery
  - Images of specimens representing both *known* and *unknown* species used to train statistical models to diagnose unknown species as novel relative to known species.

Species Identification

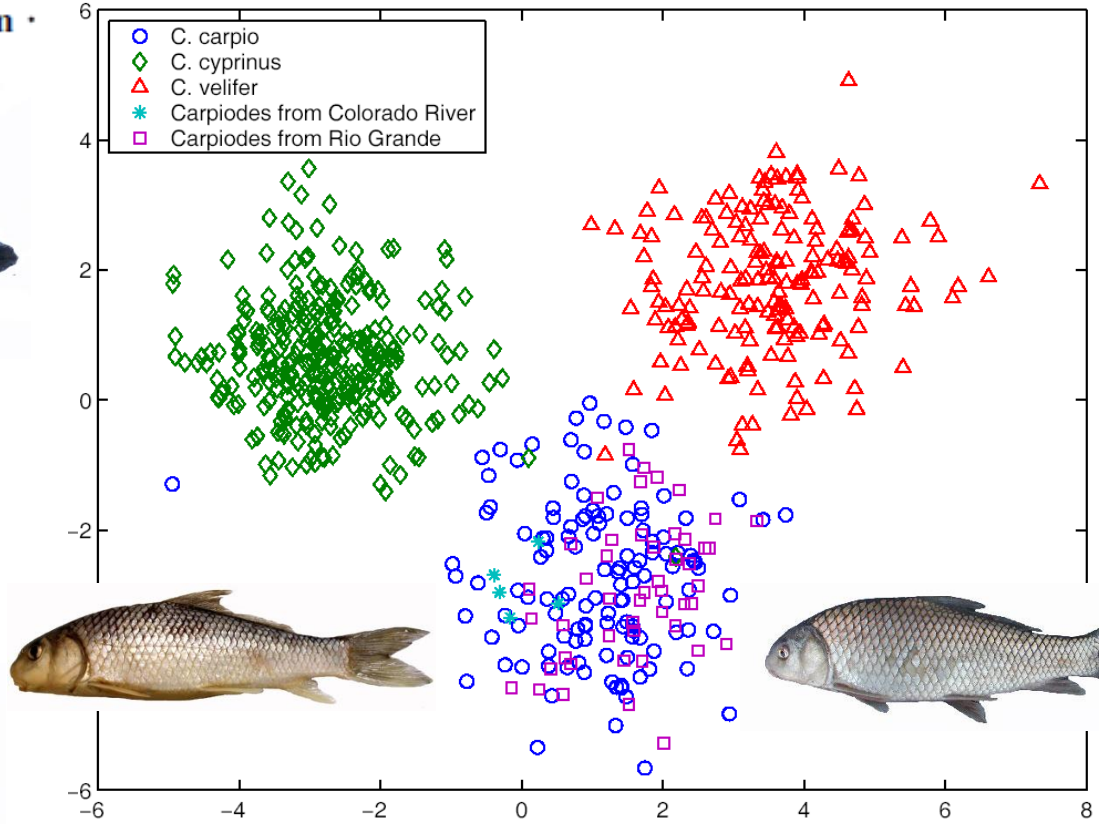
- Supervised learning:
  - Classifiers trained to predict the true species of unidentified test samples based on specimens of various known species.
  - Fundamentally assumed that test sample belongs to one of the classes that appear in the training set (i.e., is a known species).

Pattern Anal Applic (2010) 13:23–34
DOI 10.1007/s10044-009-0157-y

THEORETICAL ADVANCES

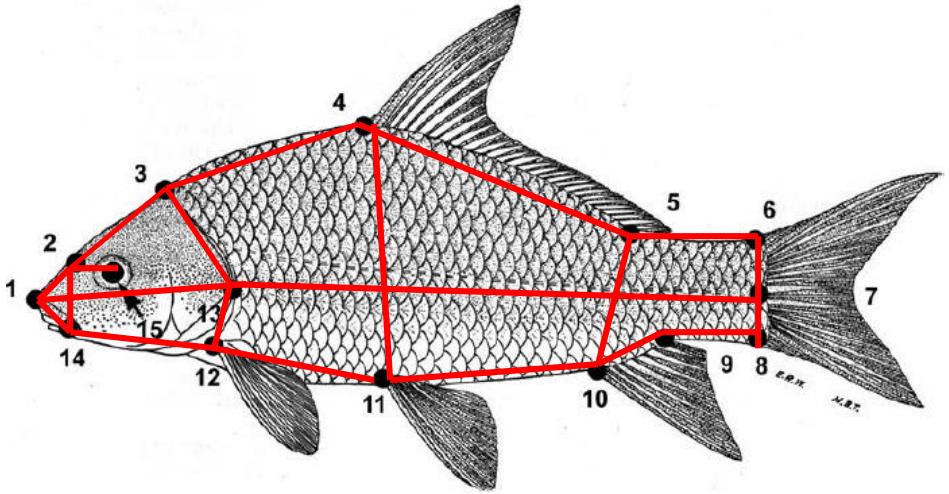
Joint feature selection and classification for taxonomic problems within fish species complexes

Yixin Chen · Shuqing Huang · Huimin Chen · Henry L. Bart Jr

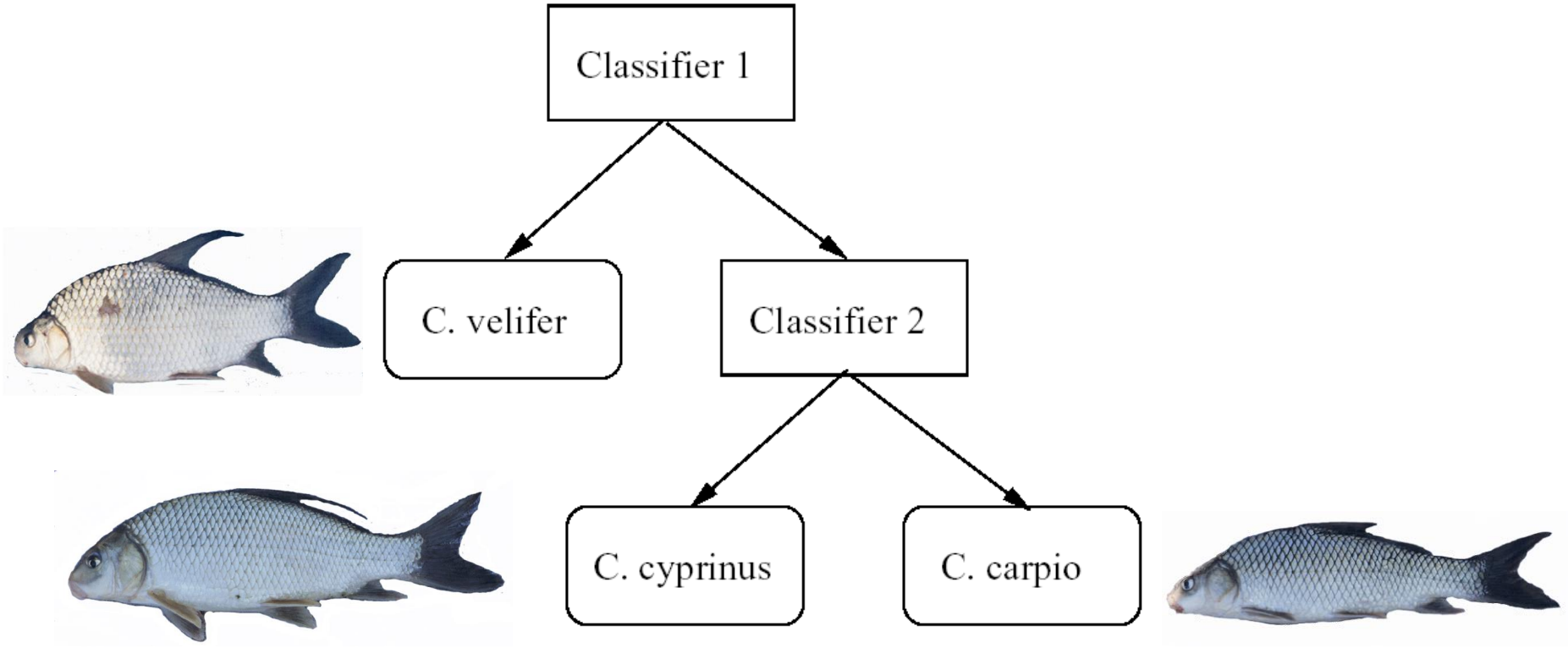


Methods

- Used Generalized Procrustes Analysis to align landmarks and remove non-shape related variation in landmark coordinates (i.e., scaled sets of landmarks of each specimen to common unit size).
- Computed 12 features, $x1, \cdots, x12$, for each specimen using inter-landmark distances and angles created at inter-landmark distance intersection points.
- Features selected; subjected to various classifiers in hierarchical, machine-learning framework.



- x_1: the distance between the tip of the snout and the naris
- x_2: the slope of the line $|LM_1\text{-}LM_2|$
- x_3: the distance between the naris and the back of the mouth
- x_4: the slope of the line $|LM_2\text{-}LM_{14}|$
- x_5: the size of head in proportion of the size of the body
- x_6: the length of the head in proportion of the length of the body
- x_7: the distance between LM_7 and LM_8
- x_8: $(|LM_1\text{-}LM_{13}|+|LM_3\text{-}LM_{13}|+|LM_{12}\text{-}LM_{13}|)/|LM_7\text{-}LM_{13}|$
- x_9: the distance between the naris and the tip of the snout in proportion of the distance between the naris and the eye
- x_{10}: the depth of the body in proportion of the length of the body
- x_{11}: $|LM_3\text{-}LM_4|/|LM_7\text{-}LM_{13}|$
- x_{12}: the slope of the line connecting LM_5 and LM_{10}



- Classifiers able to distinguish known *Carpiodes* species.
- 53 suspicious specimens were not *Carpiodes velifer* (none classified as such) and did not fit the other two known *Carpiodes* species very well (only 30-80% of specimens classify as such depending on features and classifiers used).
- Good example of species identification system, not very useful for new species discovery.

New Species Discovery

- From a machine learning perspective, detecting new species is fundamentally different from the problem of recognizing known species.
- In new species discovery, the training set cannot contain prior knowledge of an unknown species.
- Sought to develop an efficient novelty (**outlier**) detection framework using ***Statistical Depth Functions*** (SDFs).
- SDFs indicate “extremeness” of a collection of data points by providing “center-outward” ordering of multi-dimensional data.

Outlier Detection Algorithm

288 IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 31, NO. 2, FEBRUARY 2009

Outlier Detection with the Kernelized Spatial Depth Function

Yixin Chen, *Member*, *IEEE*, Xin Dang, Hanxiang Peng, and Henry L. Bart Jr.

- Training set consists of observations of known species and outliers with unknown labels.
- Kernelized Spatial Depth Function establishes threshold for distinguishing outliers from normal observations.
- Observations with depth functions less than threshold declared outliers (not part of known classes, i.e., new species).

Machine Learning in Taxonomic Research

National Science Foundation
WHERE DISCOVERIES BEGIN

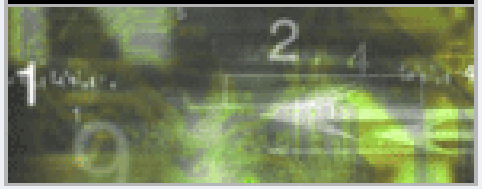
QUICK LINKS

SEARCH

HOME | FUNDING | AWARDS | DISCOVERIES | NEWS | PUBLICATIONS | STATISTICS | ABOUT NSF | FASTLANE

Email Print Share

Funding



- Find Funding
- A-Z Index of Funding Opportunities
- Recent Funding Opportunities
- Upcoming Due Dates
- Advanced Funding Search
- Interdisciplinary Research
- How to Prepare Your Proposal
- About Funding

NSF-wide

Cyber-Enabled Discovery and Innovation (CDI)

This program has been archived.

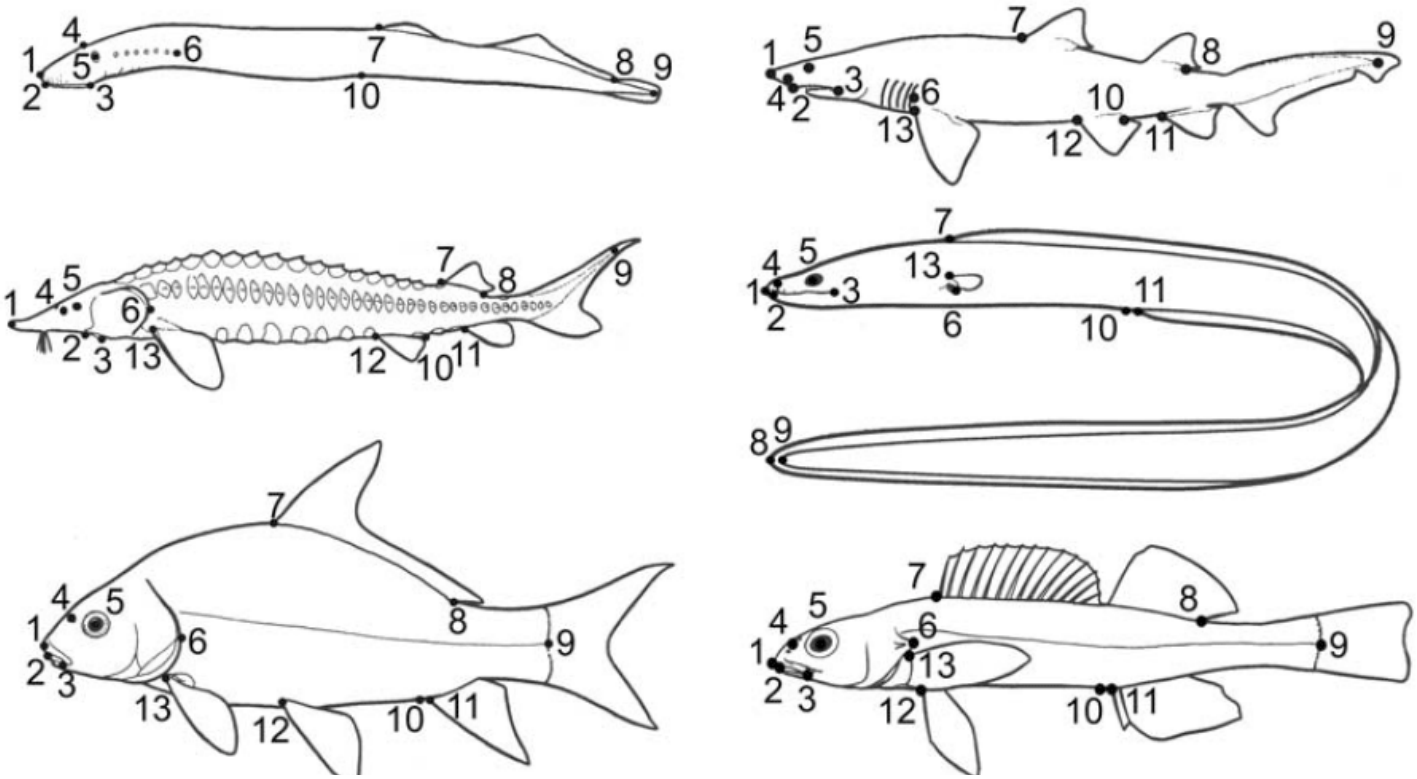
CDI Update

The Cyber-Enabled Discovery and Innovation program has demonstrated the value of interdisciplinary computational and data-enabled science and engineering. Increasingly, this research approach is being integrated into new and continuing NSF programs and solicitations. As of fiscal year 2012, proposals will no longer be accepted by the CDI program. Investigators are referred to related NSF funding opportunities, which are listed on this web page (http://www.nsf.gov/cdi). Please check for updates on this page as new opportunities are announced, and follow the links for program information and program officer contacts.

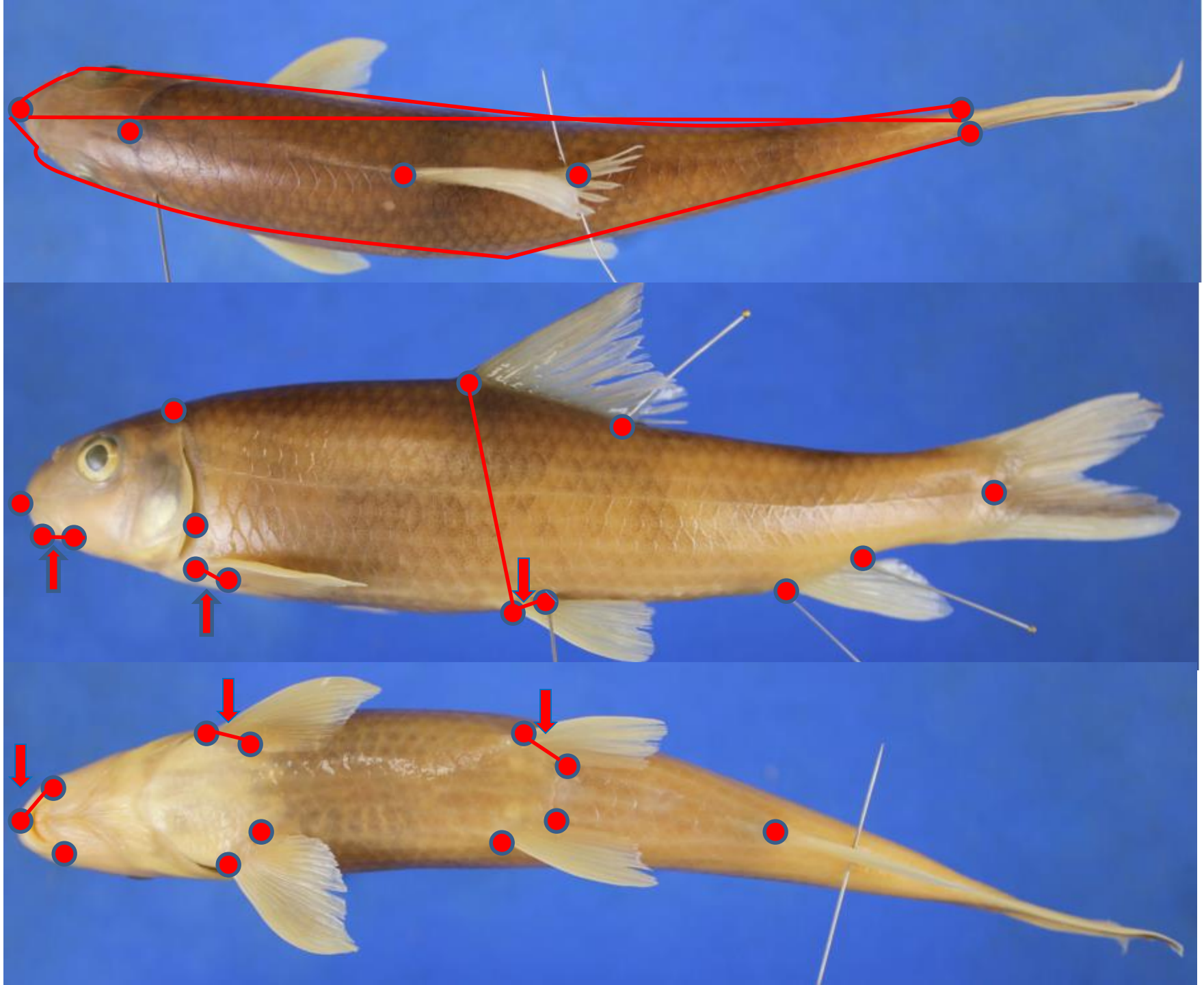
CDI Project Details

- Proposed to build an annotated database of 10,000–15,000 2D images (including **side, top and bottom views of each specimen**) corresponding to 100–150 different taxa from multiple fish families (100-150 specimens per taxon).
- Images drawn from species complexes suspected or known to contain undescribed species to prove the concept that algorithms can be trained to both identify known species and detect unknown species.
- Initially relied on body shape characters, but planned to experiment with meristic characters (e.g., squamation) and pigmentation patterns.
- Also proposed to experiment with automated landmark placement.

Identification of Higher Fish Groups



Used presence/absence of tagged landmarks and variation in homologous landmark position to diagnose families.



CDI Project Details

- Concluded that in order capture true characters of 3D fish needed to work with 3D images.
- Used NextEngine Scanner to produce 3-D surface scans of representatives of all fish families (proposed ~1000 scans) for family-level diagnosis of any fish specimen.
- Soon learned that laser scanners don’t work well with fluid-preserved specimens.

NextEngine 3-D Scanner

Specimen Preparation

- Specimen soaked in water to remove surface alcohol, then towel dried.
- Mounted in hobby vice.
- Insect pins used to identify inconspicuous anatomical features.
- "Marking Magic", a chalk-based aerosol quilting stencil spray, used to coat the specimens prior to scanning.

3-D Scanning Demo

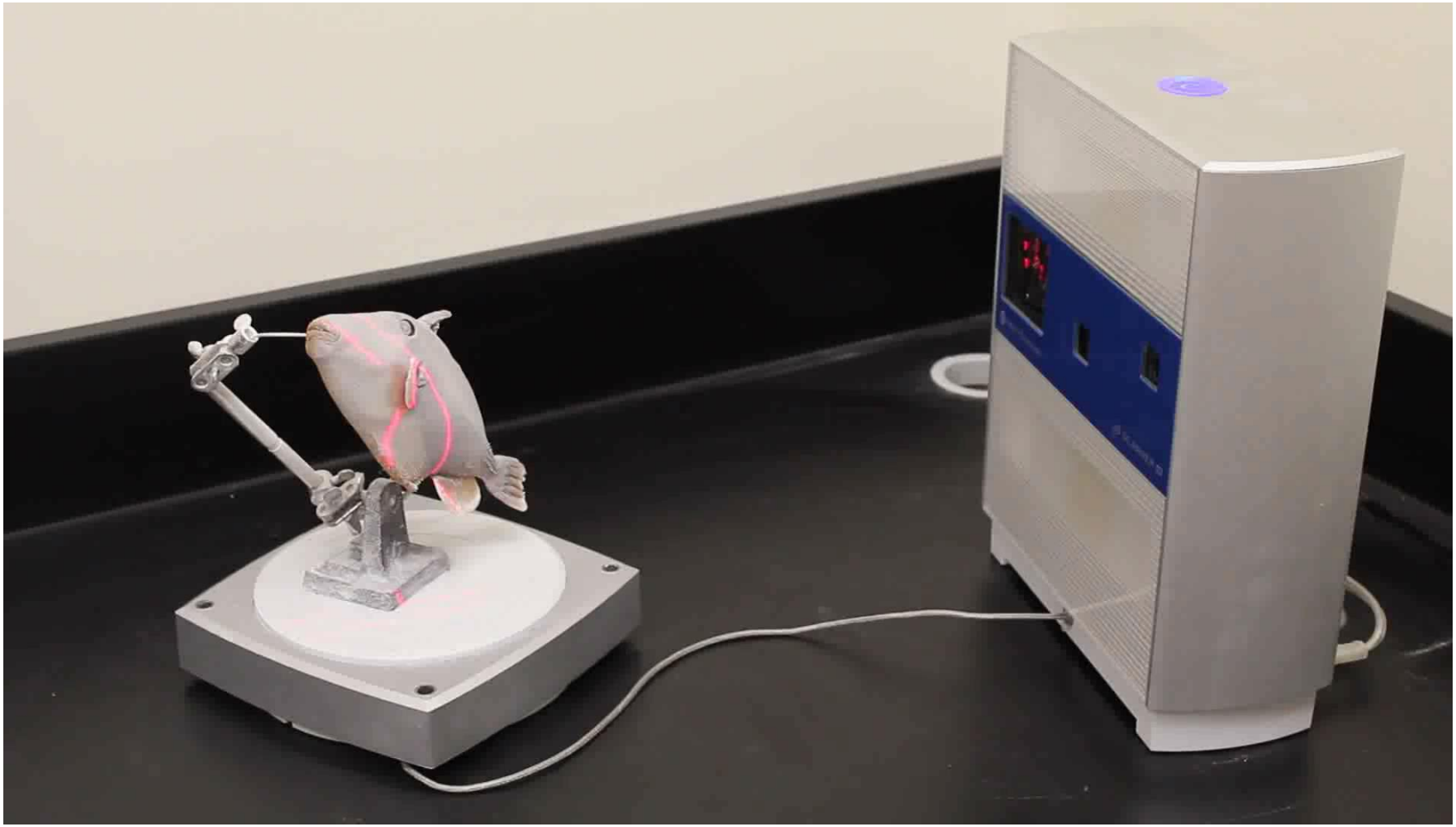
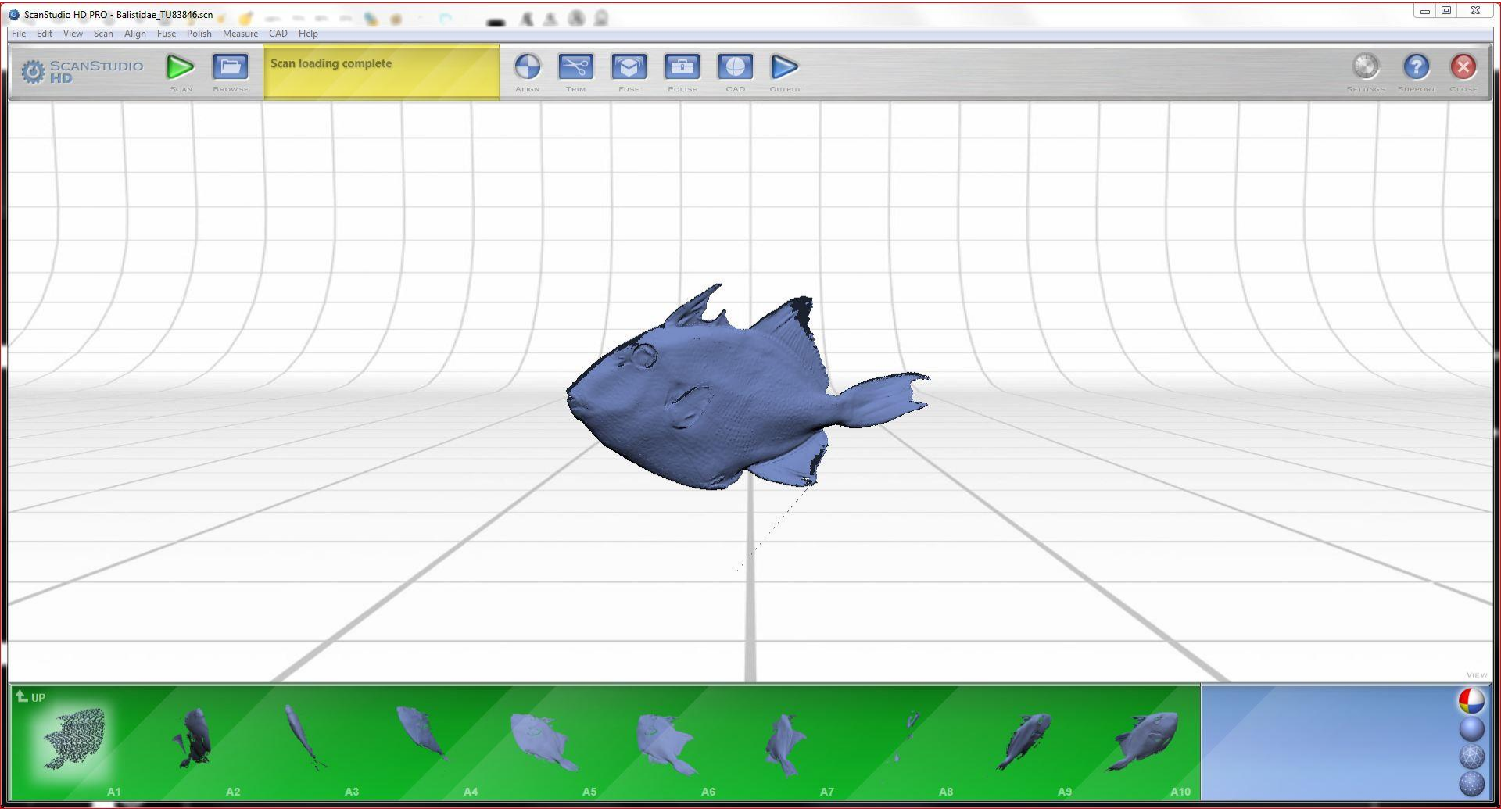
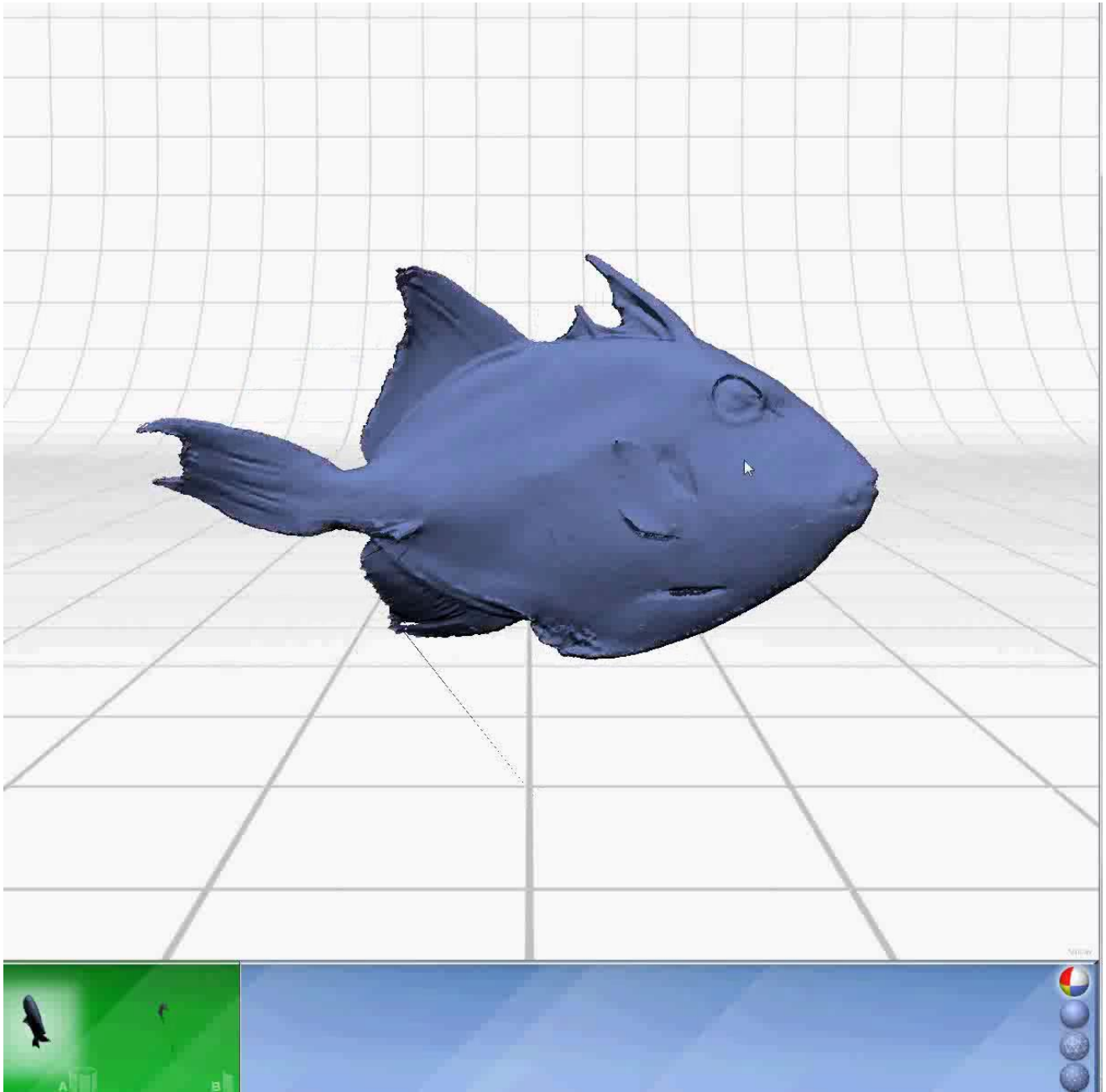
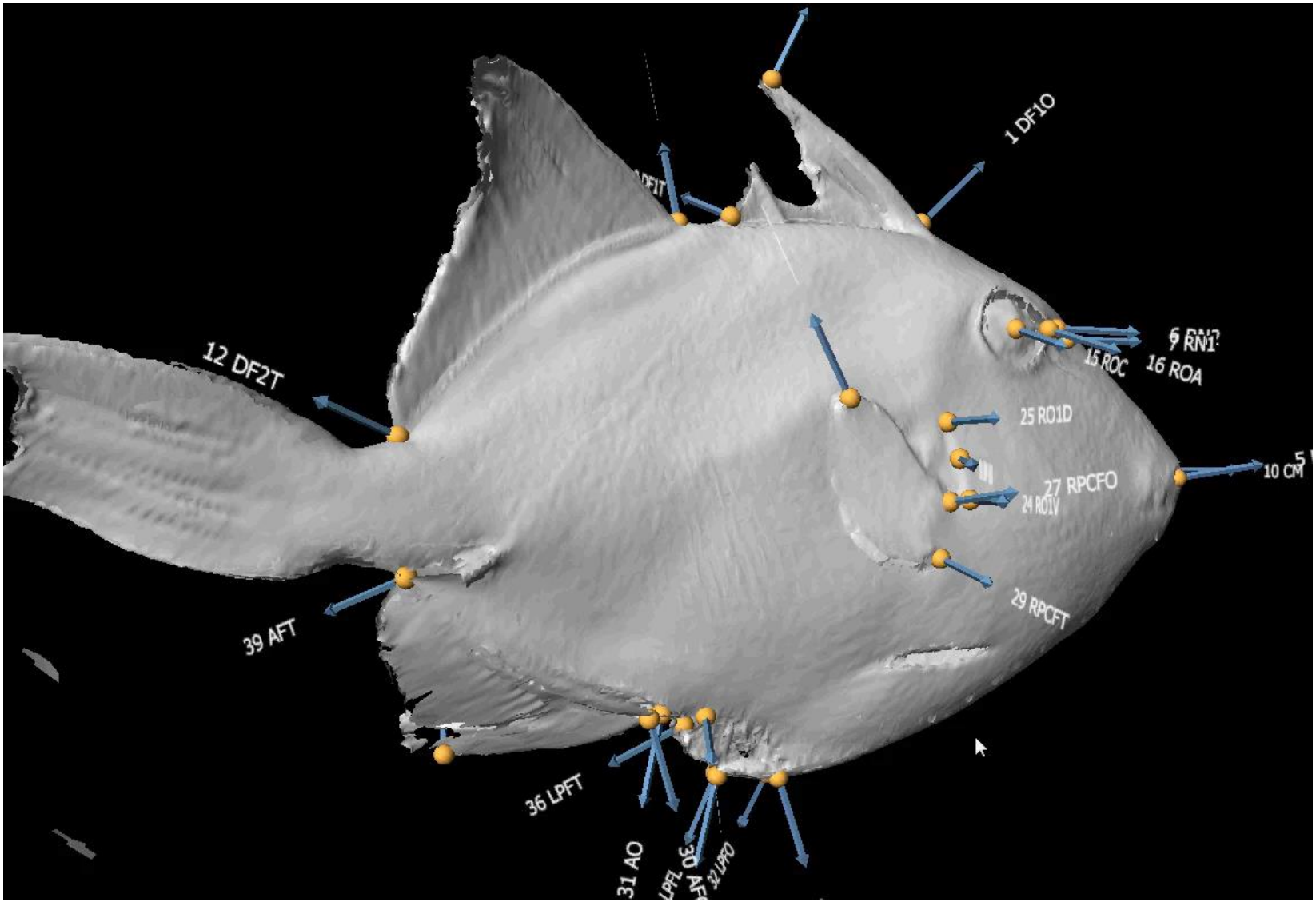


Image Stitching



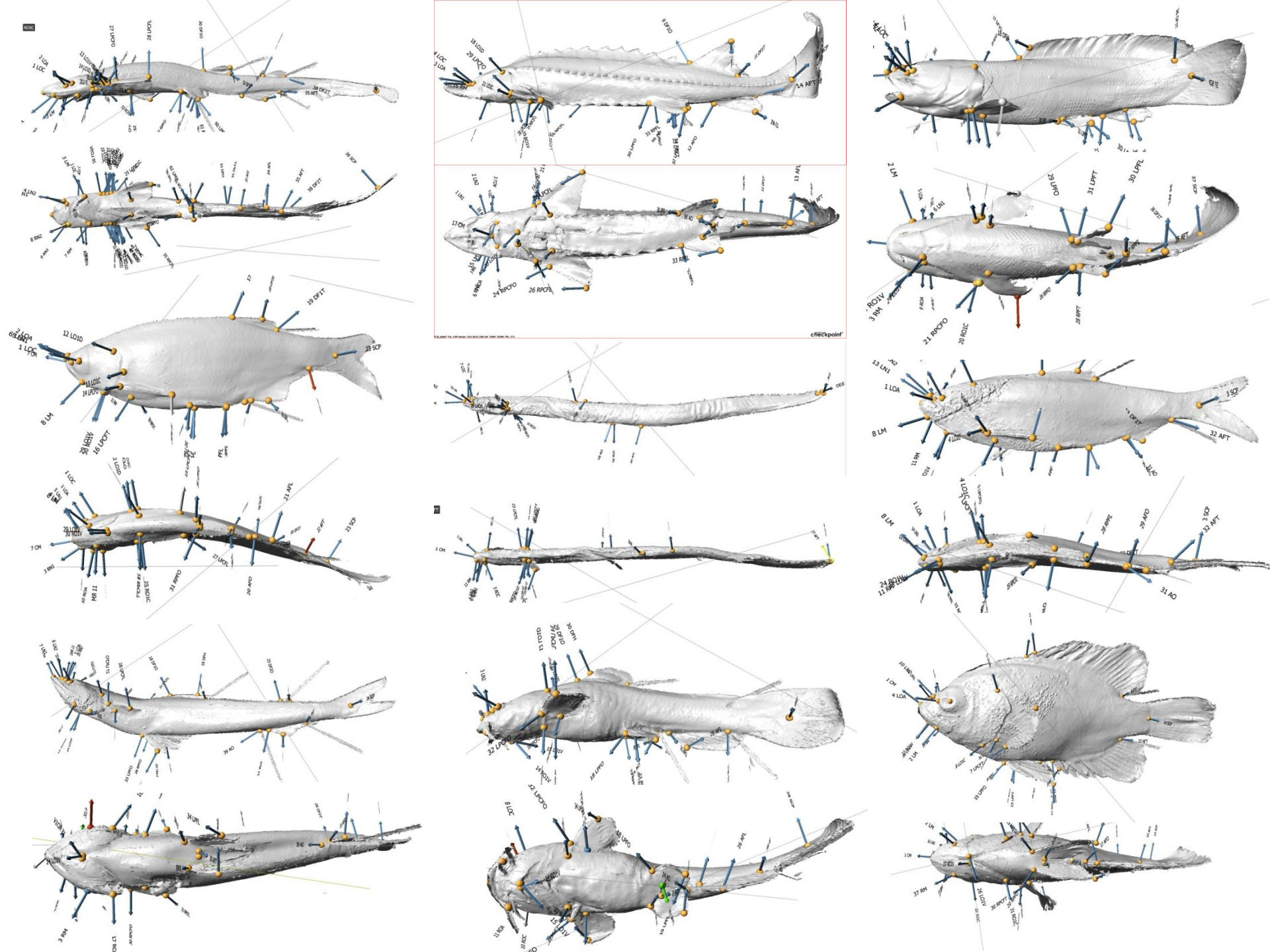


3-D Landmarking

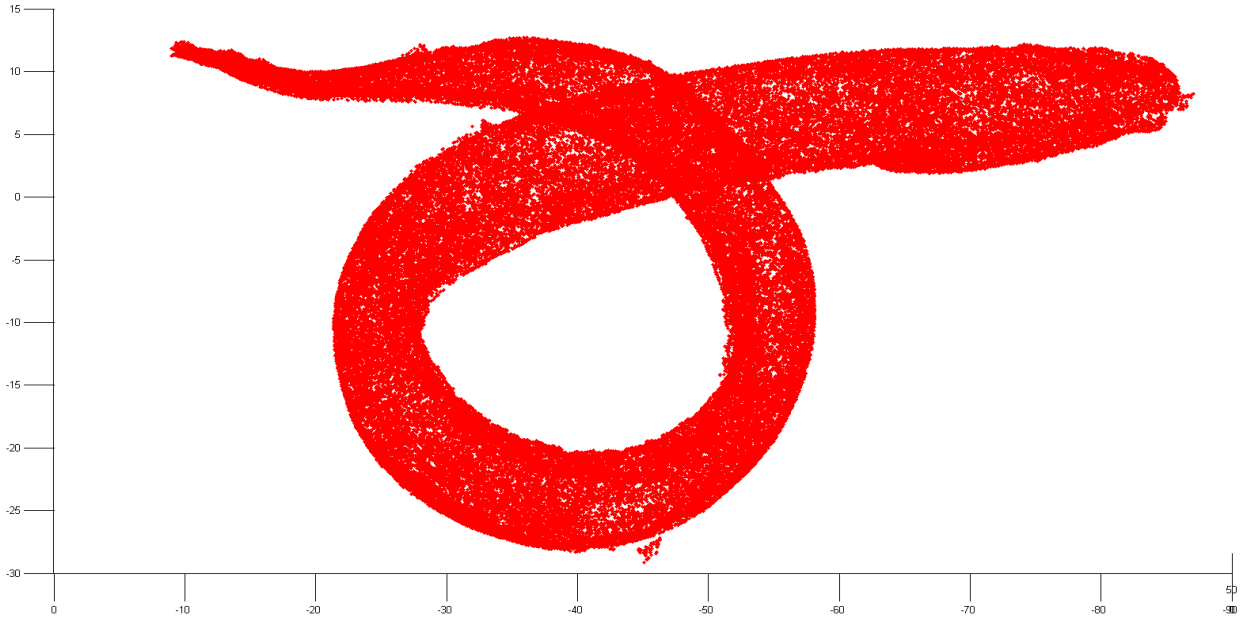


Progress of 3D scanning

- Have produced 120 3D scans representing 108 fish families (~20% of fish families).
- Have landmarked 3D scans for 89 families.
- *Would need to borrow specimens from other museums to scan additional families.*
- Have begun experimenting with feature selection for distinguishing among families and using the features to identify family membership of 2D images.



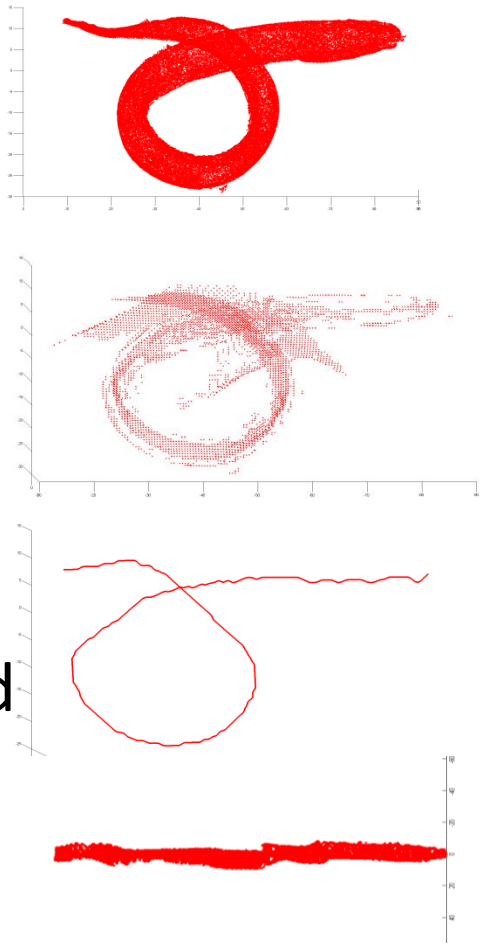
Uncurving 3D Scans of Curved Specimens



- Varying degrees of curvature is common in museum specimens.
- Doctoral student, James Church at Ole Miss, is experimenting with using Statistical Depth Function to uncurve 3D scans of curved specimens.

Uncurving 3D Scans of Curved Specimens

- Process involves sampling surface of 3D scan.
- Used statistical depth function to create a depth field of the entire scan in 3D.
- Used Dijkstra's Shortest Path algorithm to form a spine from head to tail along the path of the center of the depth field.
- Interpolated and smoothed the spine.
- Sliced the surface into sections and mapped each slice to its center spine.
- Uncurved the spine to decurve the scan.



CDI Post Mortem

- Hoped to build a portal that others could upload collections of 2D images to, and submit for family-level identification and species diagnosis.
- The system would return characters useful for rapid taxonomic identification and diagnosis.
- System's knowledge of fish diversity would increase as more species are categorized and identified.
- But none of this happened.

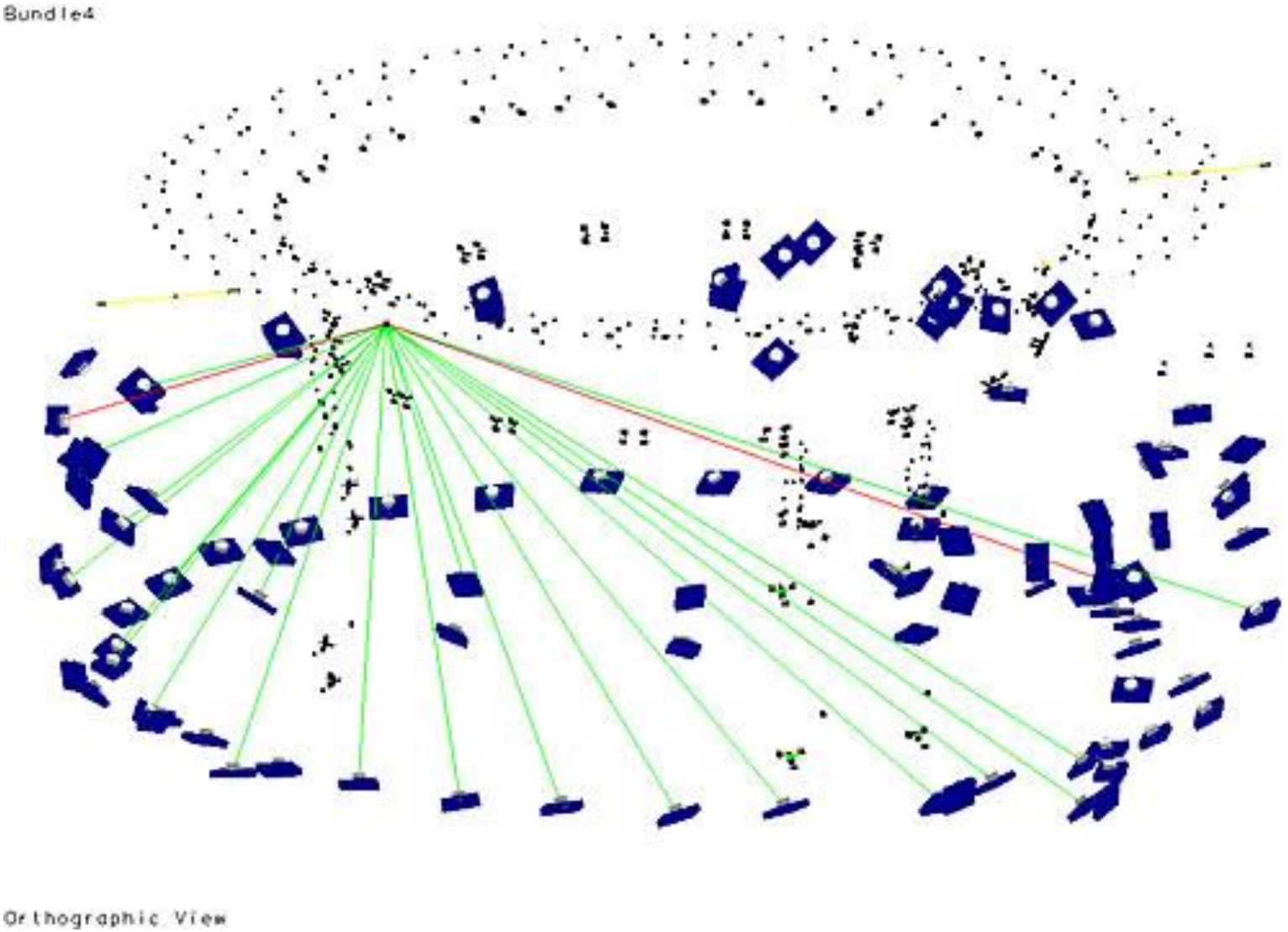
CDI Post Mortem

- Unsatisfied with data derived from 2D images.
- Sought a better way of producing 3D images than laser scanning.
- Started exploring other methods of producing 3D surfaces in 2013.
- Settled on ***photogrammetry***.

Photogrammetry

- Photogrammetry is the science of making measurements from photographs, especially useful for recovering the exact positions of points on surfaces.
- Photogrammetric 3D reconstruction involves computer processing of multiple photographic images taken around a centrally positioned object.
- Produces a 3D point cloud of the object with a surface that can preserve fine details of the object.

Photogrammetry



Photogrammetric Imaging System

- We didn’t have a system that we could use to produce 3D reconstructions of fish specimens when we first started exploring photogrammetry.
- But Nelson determined that we could do this manually…

Manual Photogrammetric Imaging System

Photogrammetry Subject

- I took 135 photos of this central stoneroller specimen (5 different planes, 360^{o}) in 20 minutes.

Photogrammetry Results

- Nelson used the images to produce this:

http://copis.tubri.org/pointsclouds/examples/stoneroller.html

Computer-Operated Photogrammetric Imaging System (COPIS)

- Photographing fluid specimens manually takes time and exposes the specimens to too much drying.
- Nelson designed and built a computer-controlled, photogrammetric imaging system employing a single mounted SLR camera.
- Still took to long to image specimens.
- Determined that we could shorten the time by increasing the number of cameras.
- Funded by IMLS Sparks! to build a 6-camera system.

Thanks for your attention!

Nelson will now discuss technical details of Photogrammetry and COPIS