TGDA + Neuroscience Project: Shape Analysis and classification via Persistent Homology

Student: Yuting Fang

Instructor: Facundo Mémoli

fang.564@osu.edu

May 28 2020

0. Problem Statement

Input: Non-Rigid World TOSCA database of 3D shapes.
About 10 classes, 150 shapes, 3000 vertices in each object.

Output: Classification between different classes

Challenges:

- Dataset too large to process all (3000*150 3D points)
- Discriminate shapes from ***different class***, without being overly sensitive to ***different poses*** of a given class.

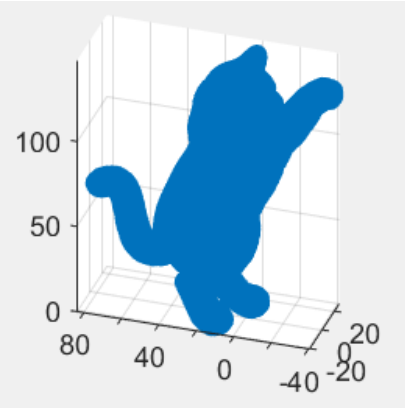
Figure 1: Different poses of the cat shape.

point clouds

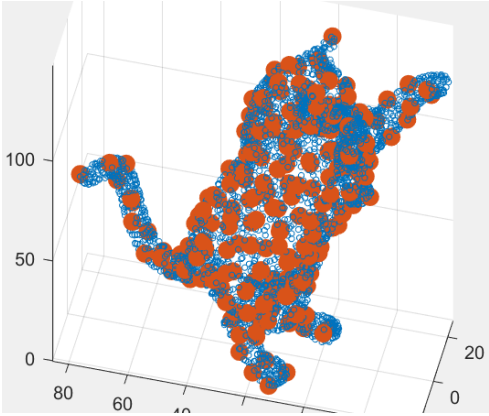
1. Basic Pipeline

Surfaces, Point Clouds → **Distance Matrix**

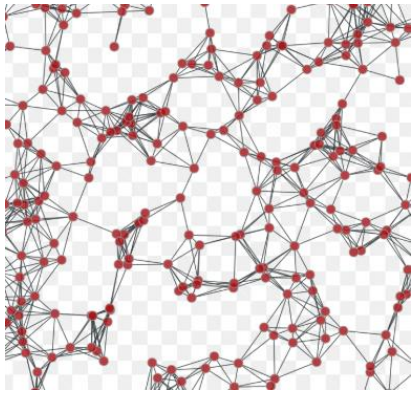
Subsample → Build graph → Compute geodesic distance



CAT1 in TOSCA



NFPS = 200
(take 200 points)



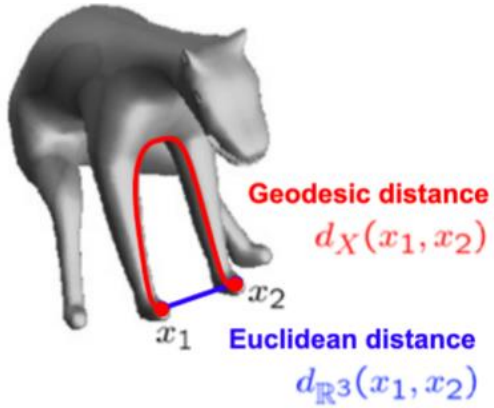
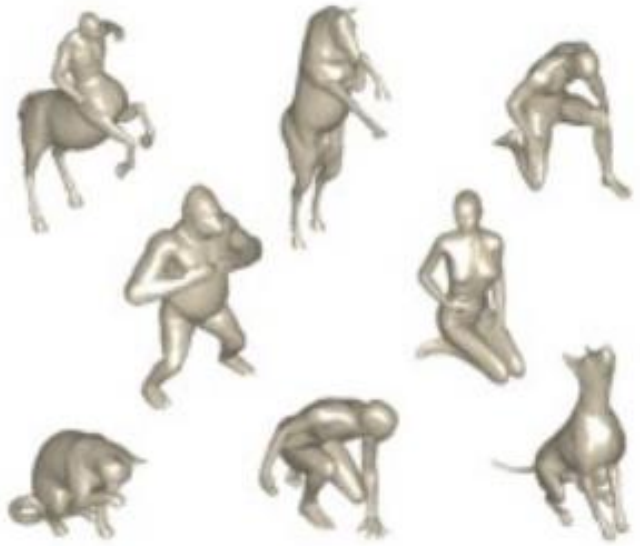


Image by Bronstein^2 and Kimmel

1. Basic Pipeline

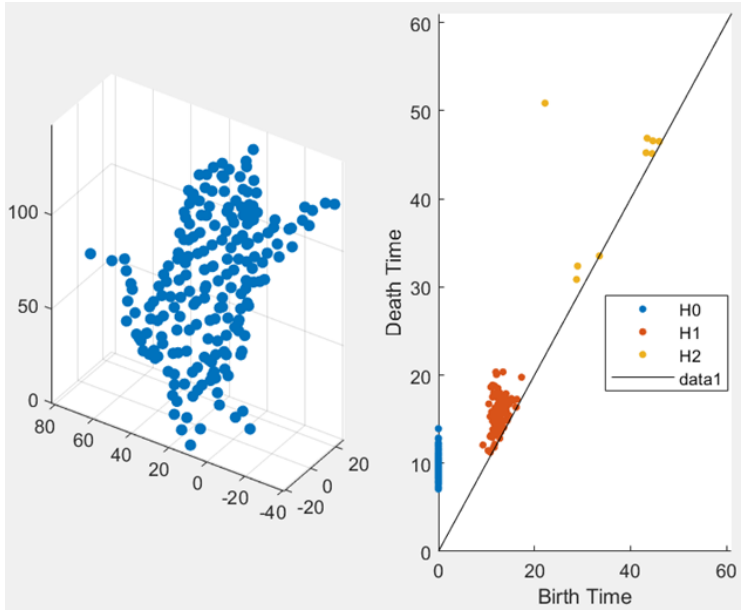
Surfaces, Point Clouds → Distance Matrix



Ripser

Do computation of Vietoris–Rips persistence barcodes [2]

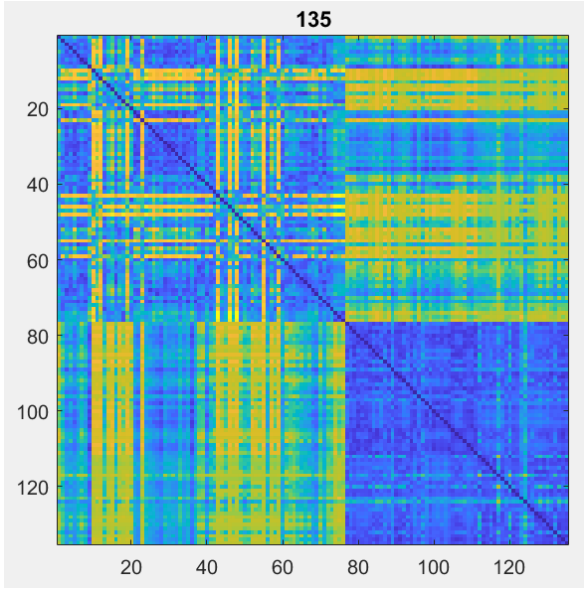
Persistence Diagrams, Barcodes



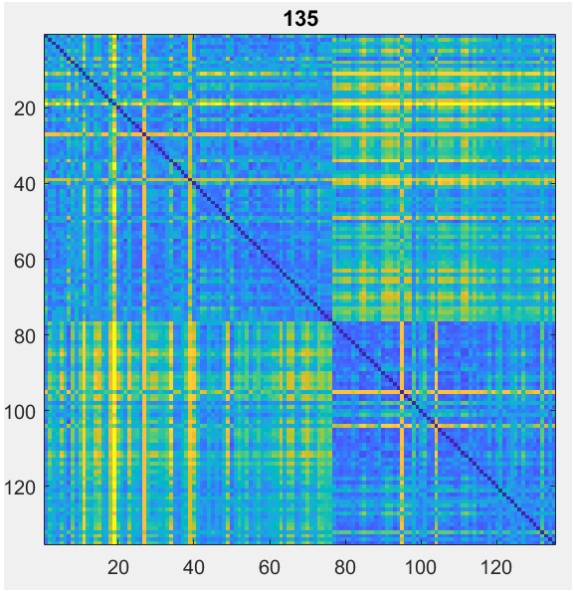
2. Bottleneck Distance Matrices

Non-Human

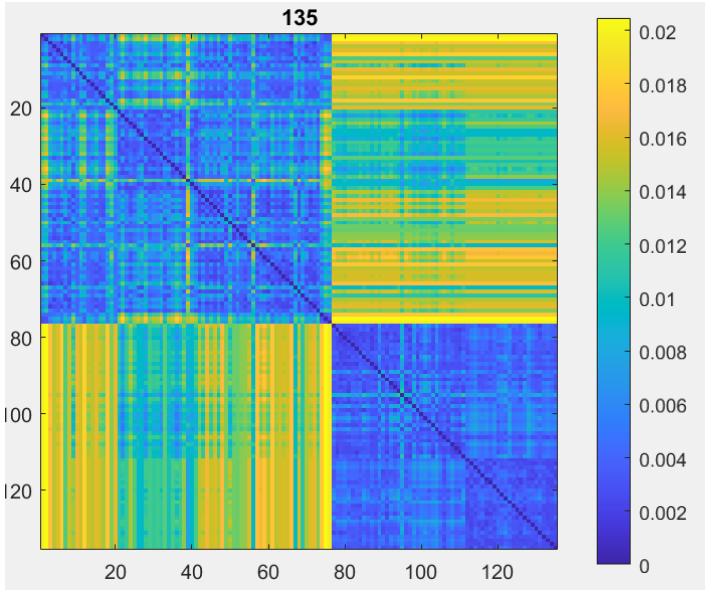
Human



PH0 dB with NFPS = 50

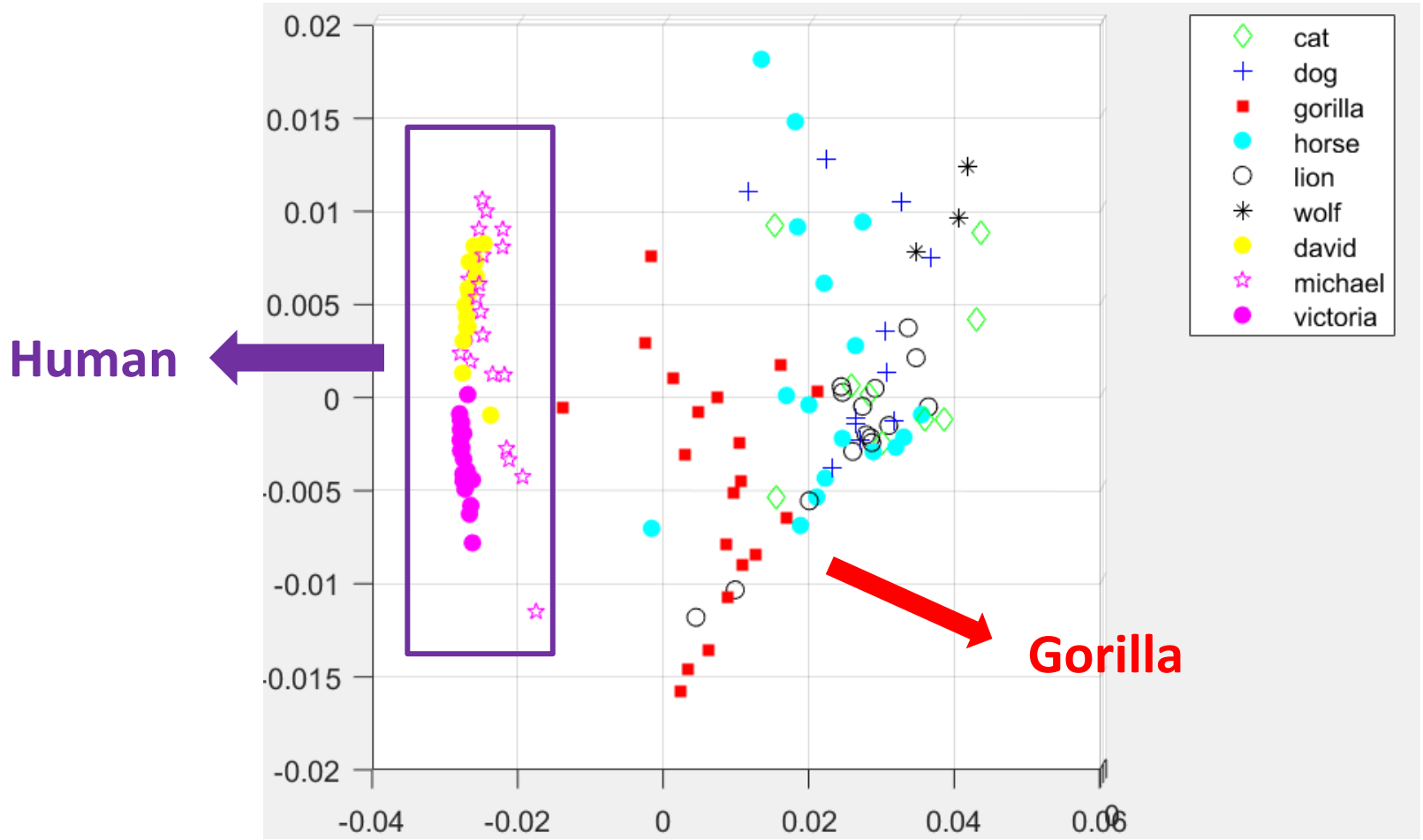


PH1 dB with NFPS = 50



PH2 dB with NFPS = 50

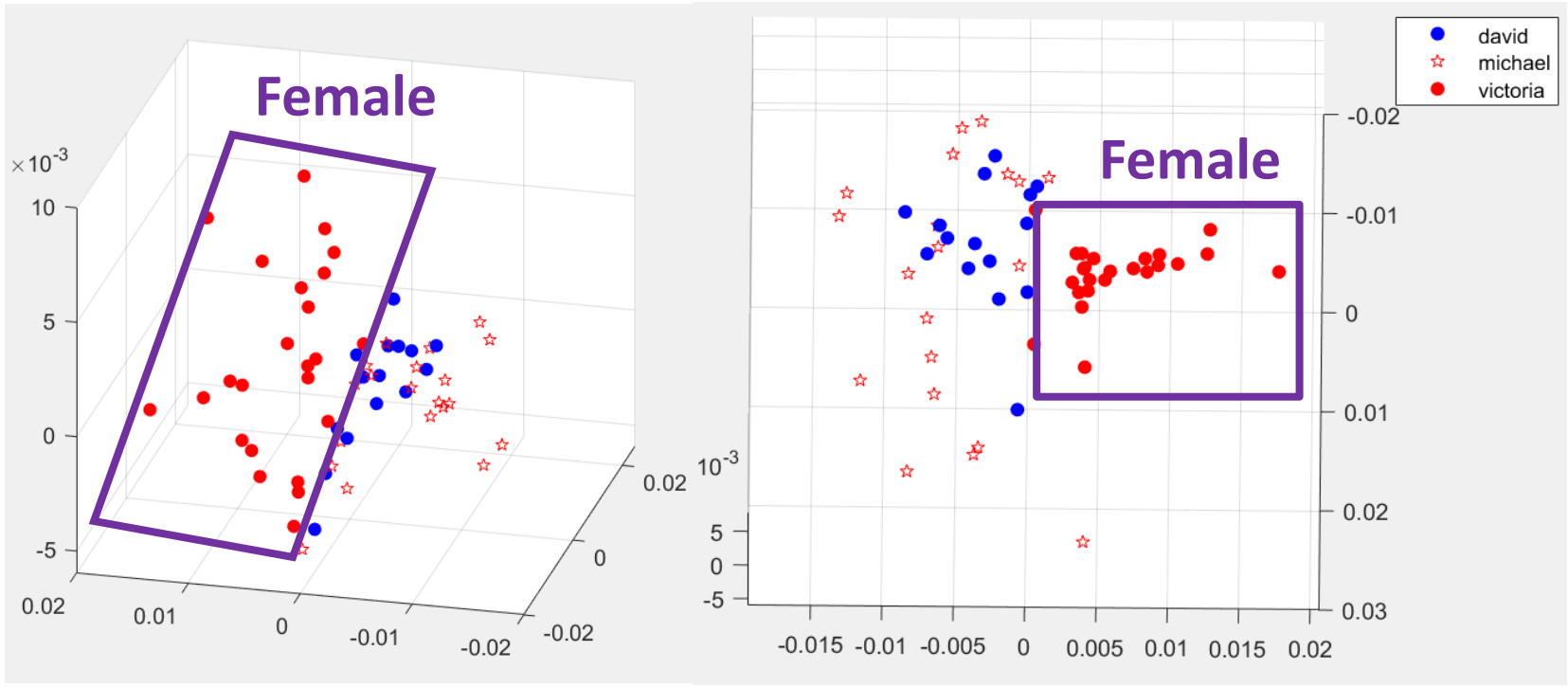
3. Visualization: Multi Dimensional Scaling



with **`cmdscale`** function in MATLAB based on PH2

3. Multi Dimensional Scaling - Human

- Classify well: male vs. female
- Problem: male vs. male ?



Actually Twins

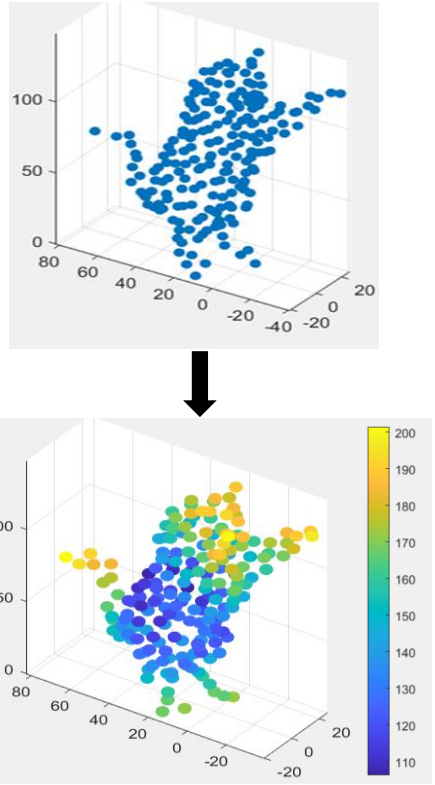
Michael15 David2

4. Next Step

- Apply Machine Learning to get better model
  - Better combination: $A = A_0 \times dB0 + A_1 \times dB1 + A_2 \times dB2$
  - k-Nearest Neighbors : `Mdl = fitcknn( )`

- Apply **Eccentricity Function** to ***enlarge features*** *[1]*

  For two connected points x, y: $G(x, y) = 1$
  - Current Weight: $w(x, y) = ||x - y||$
  - Better Weight: $w(x, y) = ||x - y||\boldsymbol{e(x)e(y)}$
  - Normalization



What I learn these days

- Experience of **processing large geometric database**
  - **Subsampling**: random, FPS, …
  - **Noise Canceling**: which points to choose
  - What I should prepare for: huge dataset, long running time…

- Experience of **Topology Data Analysis**
  - Basic ideas of **Persistent Homology**
  - Important features and their meaning: **DM, PD, Barcodes**, …

Reference

[1] *Gromov-Hausdorff Stable Signatures for Shapes using Persistence*, Frédéric Chazal, 2009.

[2] *Ripser: efficient computation of Vietoris-Rips persistence barcodes*, Ulrich Bauer, 2019.