Texture Based Classification of Seismic Image Patches Using Topological Data Analysis

Brad Nelson

Institute for Computational and Mathematical Engineering,
Stanford Linear Accelerator,
Stanford University

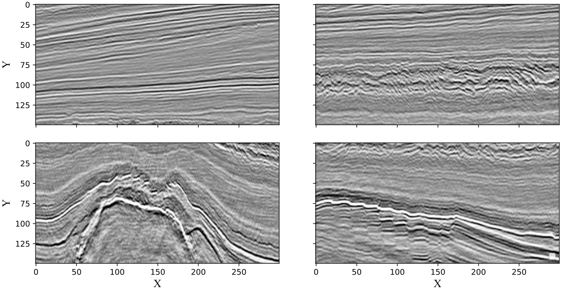
SEP Group Meeting
May 3, 2019

Acknowledgments

All content in this talk is joint work with Rahul Sarkar.

Biondo Biondi and Gunnar Carlsson (our advisors) gave lots of useful guidance along the way.

Seismic Texture Classification

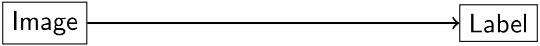


→ Flat Horizon
Chaotic Horizon
Fault
Salt Body

Example image patches from the LANDMASS-2 data set [4].

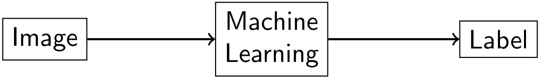
This talk: one way to do this using topological data analysis.

Seismic Texture Classification



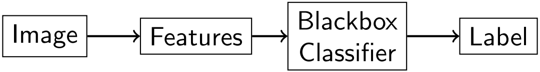
What we want

Seismic Texture Classification



A popular approach.

Seismic Texture Classification



Now we have a road map

Texture

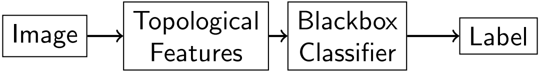
Characteristics of texture

- Area with similar 'look' - no *one* thing to focus on
- Repetitive/recurrent (but not necessarily periodic)
- Scale, translation, rotation, deformation invariant

Machine learning with images

- Treating each pixel as a feature is typically not well-suited for classification tasks.
- Most existing methods use derived features (either hand-crafted or learned).
- We'll show how topological features capture the properties of texture.

Seismic Texture Classification



Where we're headed

Topology

Topology:

- Scale, translation, rotation, deformation invariant
- Good at capturing qualitative information

Topological Data Analysis:

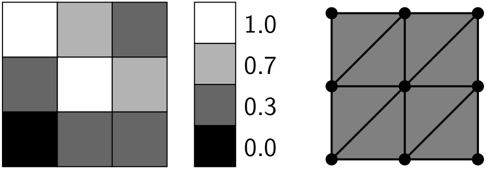
- Generally: tools to use and understand topological structure in data
- Tools to turn topology into features (real numbers)

Topology in Images

To use topology on images, we need a way to turn an image into a topological space. We represent the space on a computer using points, edges, and triangles (simplicial complex).

One way to do this:

- Pixels become points in the space
- Adjacent pixels are connected by an edge
- Diagonal edges added by Freudenthal triangulation
- 3 adjacent pixels are spanned by a triangle

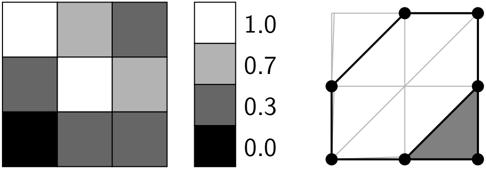


Topology in Images

Topological spaces created from all pixels in the image have the same topology as the ‘canvas’ (e.g. rectangle). This won’t help with classifying different images.

A more interesting topological space:

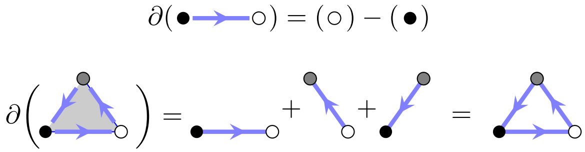
- Choose some pixel value t
- Only points with pixel values are $\leq t$ are used
- Only edges with both endpoints are included
- Only triangles with boundary edges are included



$t = 0.7$

Homology

Consider formal linear combinations of vertices/edges/triangles in a space X. This produces a set of vector spaces $C_k(X)$ ($k = 0$ for vertices, $k = 1$ for edges...). There are linear boundary maps $\partial_k : C_k(X) \to C_{k-1}(X)$



with the property $\partial \circ \partial = 0$.
Homology of the space in dimension k is the quotient

$$H_k(X) = \ker(\partial_k) / \operatorname{img}(\partial_{k+1})$$

$\dim H_0$ counts clusters that are not connected. $\dim H_1$ counts cycles (holes) that are not boundaries.

Filtrations

A filtration is

- a set of spaces X_t, $t \in I \subseteq \mathbb{R}$
- the spaces satisfy $X_s \subseteq X_t$ if $s \leq t$

We obtain a filtration from an image by constructing spaces X_t for every pixel value t in the image.

X_t is the space constructed from pixel values $\leq t$.

Persistent Homology

Persistent homology is a tool from topological data analysis

Input: A filtration X_t

Output: A collection of pairs of real numbers for each dimension k in the space

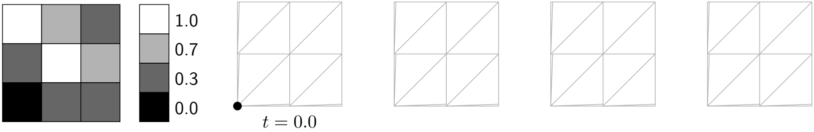
$$PH_k(X_t) = \{(b_j, d_j)\}$$

These are called birth-death pairs, and track how homology changes over the filtration.

Properties:

- homotopy invariant (deformation, rotation, translation)
- stable to perturbations of pixel values

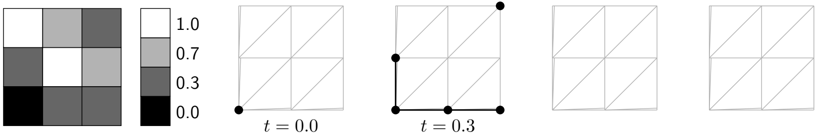
Example



Left: Example image Right: Corresponding filtration

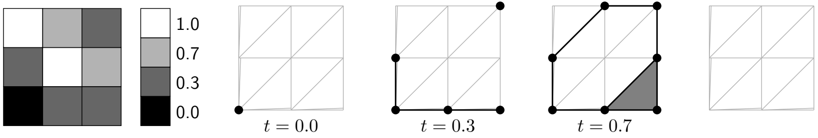
At $t = 0$, a single point appears...

Example



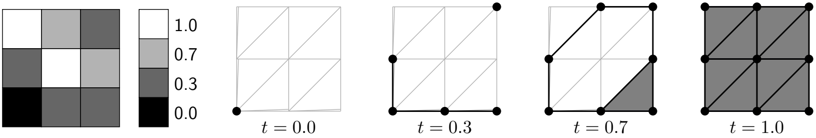
At $t = 0.3$, several points connect to the first point, and a new component emerges

Example



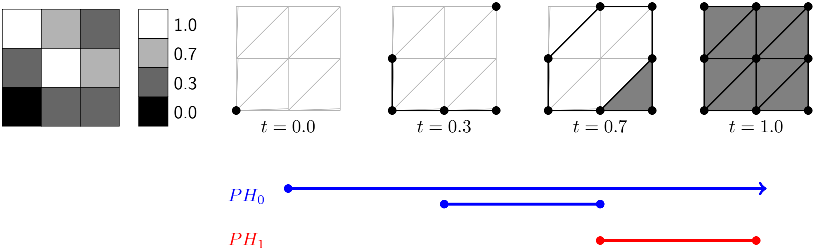
At $t = 0.7$, the two components join, and a hole appears.
We also see our first triangle.

Example



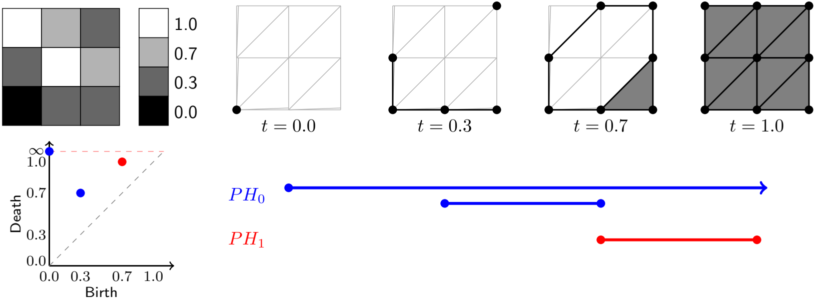
At $t = 1.0$, all points are now present, and all edges and triangles fill in the space

Example



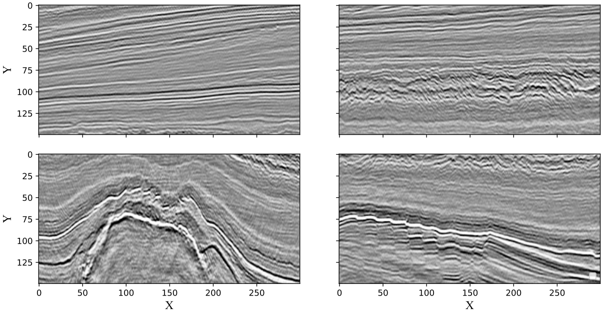
Persistence Barcode:
Information about how components appear and merge is encoded in PH_0
Information about how holes appear and fill in is encoded in PH_1

Example

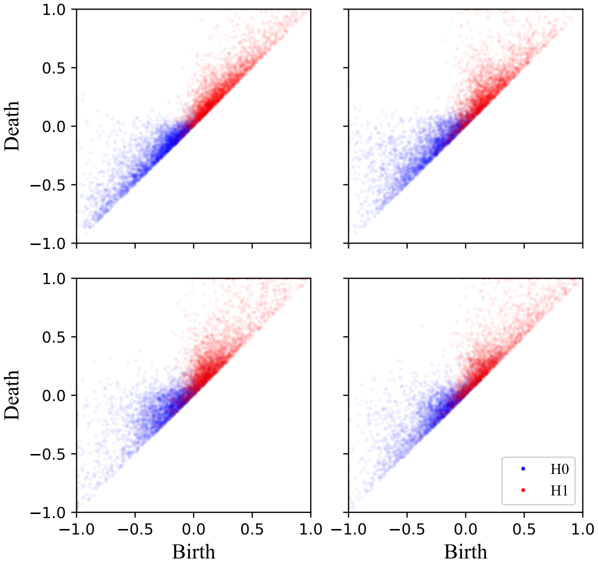


Persistence Diagram:
The start and endpoints of the barcode are plotted in the plane.
Each point is referred to as a birth-death pair.

Seismic Texture Classification



Seismic Texture Classification



Features

We went from an image to a set of 2D points

- Every image produces a different number of birth-death pairs
- We want a standard number of features

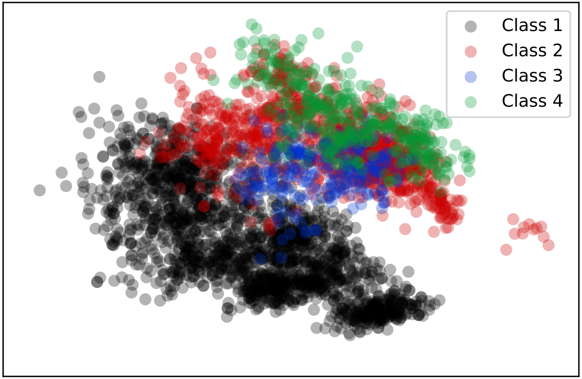
One approach - polynomial functions [2]:

$$p(\alpha; \{(b_i, d_i)\}_{i \in J}) = \frac{1}{|J|} \sum_{i \in J} \sum_{j,k} \alpha_{j,k} (d_i - b_i)^j (d_i + b_i)^k$$

Choose several different α, and we have as many features. We use $\alpha = \delta_{ij}, (i, j) \in \{0, 1, 2, 3\}^2 - (0, 0)$ for homology dimensions 0 and 1. This produces $15 \times 2 = 30$ features.

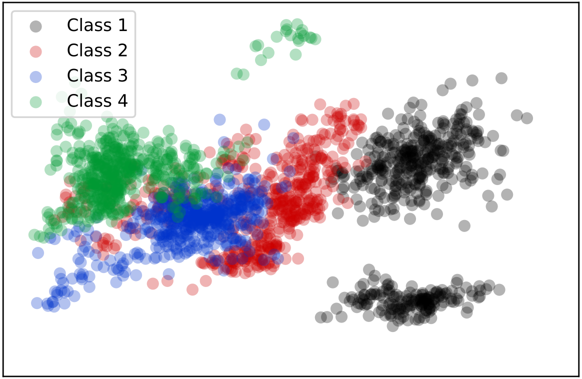
These look a lot like empirical expectations of functions using point measures in $\mathbb{R}^2$.

LANDMASS-1 Features



Projection of topological features into top 2 principal components. Each point is an image in the LANDMASS-1 data set.

LANDMASS-2 Features



Projection of topological features into top 2 principal components. Each point is an image in the LANDMASS-2 data set.

GLCM Features

We also investigate the use of grey level co-occurence matrices (GLCM) features to produce new images.

Each GLCM feature produces a new image, where pixel values depend on a small patch.

GLCM Features: Entropy, Energy, Variance, Correlation, Mean, Max Probability...

We use these additional GLCM images to produce additional persistence diagrams for each input image, and use the same polynomial features.

Method

- Split data into train and test
- Produce persistence diagrams for each image
- Produce polynomial features from each persistence diagram
- Train classifier on polynomial features
- 3 Black box classifiers tested
  - Multi-class SVM
  - Random Forest
  - Neural network

Persistent homology calculations done with GUDHI [5] SVM and RF use scikit-learn [6] NN uses tensorflow [7].

Results

Attribute	SVM Accuracy
Raw Image	99.8 / 75.2 / 0.0 / 0.0 100.0 / 55.0 / 88.3 / 74.3
GLCM Mean	100.0 / 18.6 / 34.1 / 29.3 62.7 / 19.0 / 4.0 / 100.0
RMS Amplitude	100.0 / 1.0 / 0.0 / 0.0 74.7 / 85.7 / 71.3 / 61.7
GLCM Correlation	100.0 / 0.0 / 0.0 / 0.0 64.7 / 32.0 / 89.3 / 32.3
GLCM Variance	96.6 / 94.1 / 92.8 / 67.7 97.3 / 93.3 / 91.7 / 87.0

In each row: top = LANDMASS-1, bottom = LANDMASS-2,
color corresponds to class

Results

Attribute	RF Accuracy
Raw	99.9 / 98.6 / 95.2 / 93.3
Image	100.0 / 98.0 / 100.0 / 96.3
GLCM	99.9 / 97.9 / 82.1 / 93.3
Mean	100.0 / 97.0 / 97.3 / 91.7
RMS	99.3 / 96.1 / 88.0 / 82.0
Amplitude	99.7 / 96.0 / 96.0 / 91.7
GLCM	99.3 / 94.9 / 80.8 / 91.2
Correlation	99.7 / 93.7 / 92.0 / 97.0
GLCM	98.5 / 95.7 / 96.3 / 74.0
Variance	99.0 / 95.3 / 96.7 / 89.7

In each row: top = LANDMASS-1, bottom = LANDMASS-2, color corresponds to class

Results

Attribute	NN Accuracy
Raw Image	100.0 / 99.6 / 99.7 / 98.4 100.0 / 100.0 / 99.0 / 95.0
GLCM Mean	100.0 / 97.8 / 92.8 / 97.0 100.0 / 96.0 / 95.7 / 96.3
RMS Amplitude	99.5 / 99.1 / 96.3 / 91.5 99.7 / 99.0 / 93.7 / 91.3
GLCM Correlation	99.8 / 93.6 / 87.7 / 96.7 100.0 / 95.7 / 93.7 / 98.3
GLCM Variance	99.3 / 98.3 / 98.1 / 87.3 99.7 / 99.0 / 99.3 / 95.0

In each row: top = LANDMASS-1, bottom = LANDMASS-2,
color corresponds to class

Conclusions

- Topological features potentially give high accuracy
- SVM had poorest performance. RF and NN did much better, probably due to nonlinear decision boundaries
- some (not all) GLCM features also do well when combined with topological features

Some next steps

- Analysis of space of topological features (not just classification accuracy)
- How many polynomial features do we need?
- (Further out) 3D images

References I

- R. Sarkar, B. Nelson. Texture Based Classification of Seismic Image Patches Using Topological Data Analysis. EAGE 2019 (to appear).
- A. Adcock, E. Carlsson, G. Carlsson. The ring of algebraic functions on persistence bar codes. Homology, Homotopy and Applications. 18(1) 2016.
- H. Di, D. Gao. Improved seismic texture analysis based on nonlinear gray-level transformation. SEG Technical Program Expanded Abstracts 2016.
- Y. Alaudah, Z. Wang, Z. Long, G. AlRegib. LANDMASS Seismic Dataset. `http://cegp.ece.gatech.edu/codedata/LANDMASS/index.html` 2015.
- C. Maria. Filtered Complexes. GUDHI User and Reference Manual. `http://gudhi.gforge.inria.fr/doc/latest/group_simplex_tree.html` 2015.
- F. Pedregosa et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12. 2011.
- M. Abadi et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Whitepaper. `https://www.tensorflow.org/` 2015.