This article has been accepted for publication in the expanded abstracts of the 81st Annual International Conference and Exhibition, EAGE

Texture Based Classification Of Seismic Image Patches Using Topological Data Analysis

Rahul Sarkar, and Bradley J. Nelson [1]

ABSTRACT

In seismic imaging, a long sought after goal has been either full or partial automation of the seismic image segmentation and interpretation processes. In this study, we present a novel supervised learning method for textural classification of seismic image patches, based on a topological tool called *persistent homology*. The basic workflow starts by taking an image and calculating its persistent homology, which gives us a list of birth-death pairs for different homology dimensions. Polynomial feature vectors are then extracted from these pairs, which are used to train three commonly used classifiers — support vector machines, random forests, and neural networks, whose performances we compare. In addition, we experiment with different derived textural attributes and test the impact of using them instead of the raw images in the workflow. Our proposed method is tested on the publicly available LANDMASS datasets, which contains two sets of 2D seismic image patches grouped into four classes. The results indicate that persistent homology derived features can be powerful for automated textural segmentation of seismic images.

INTRODUCTION

The task of seismic image segmentation is routine in many different areas of hydrocarbon exploration, from detection of faults and salt domes, to estimating the reserve sizes. This is an extremely human intensive process, and considerable effort has gone into its automation. Early efforts towards this goal focused on developing attributes to aid human interpreters (Chen and Sidney, 1997; Chopra and Marfurt, 2005), but this still required considerable human input. More recently, with advances in machine learning (ML), strides have been made towards partial to full automation of such tasks. Supervised and unsupervised learning algorithms have been employed for seismic facies recognition and structure labeling (Zhao et al., 2015), whereas more

[1]Email: bradnelson@stanford.edu

specialized deep learning techniques have been used for automated seismic interpretation and salt body identification (Waldeland et al., 2018) tasks. A subclass of these methods attempts to perform image segmentation using textural attributes (Love and Simaan, 1984; Chopra and Alexeev, 2006), based on the observation that different lithologies often exhibit different image textures. For example, a salt body looks very different from a region of sedimentary deposits in a seismic image. An important recent work by Chevitarese et al. (2018) demonstrates that seismic image segmentation can be performed by breaking up the image into small patches, followed by using convolutional neural networks (CNNs) to classify these patches based on their textures.

In this abstract we demonstrate the use of features obtained from topological data analysis (TDA) (Carlsson, 2009) for seismic image classification, which are robust to a variety of challenges in texture recognition such as image rotation, scaling, and nonlinear deformation, as well as perturbations to pixel values. We use features from raw input images, and also experiment with various commonly used textural attributes as inputs to our workflow. The strength of these features is tested by training three different classifiers on two publicly available labeled 2D seismic datasets.

TOPOLOGICAL DATA ANALYSIS

In this section, we give a minimal introduction to several tools in TDA useful for our purposes; for a more thorough introduction we defer to Carlsson (2009), and Ghrist (2017). Images give rise to topological spaces in a straightforward way. We denote a space as X, in which pixels become points in the space, adjacent points are connected by edges, and minimal triangles are spanned by a 2-dimensional face. We assign each pixel its coordinates in the image, and use the Freudenthal triangulation of the integer lattice (Freudenthal, 1942). Interesting information appears when we threshold the space by pixel value. We denote X_t as the space that contains only points whose pixel intensity value is less than t, only edges whose both endpoints are included, and only triangles for which all boundary edges are included. The sequence of spaces resulting from varying t is called a (sub-level set) filtration on X.

Homology is a homotopy invariant of topological spaces, and persistent homology tracks how homology changes in a filtration. The input to persistent homology is a filtration on a space, and the output is a collection of pairs of real numbers, called birth-death pairs, in different dimensions. Figure 1 contains a working example of how persistent homology is calculated from the sub-level set filtration of a 3×3 image patch. In our situation, only two dimensions contain information. Zero dimensional information, $PH_0(X_t)$, describes connected components, and a birth death pair $(b_i, d_i) \in PH_0$ represents a connected component that appears in the filtration at $t = b_i$, and merges with an older component at $t = d_i$. There is always a single component that survives to the end of the filtration, and we say its death is at infinity. One-dimensional information, $PH_1(X_t)$, contains information about holes in the space which are not filled in by triangles. For $(b_i, d_i) \in PH_1(X_t)$, the birth b_i is

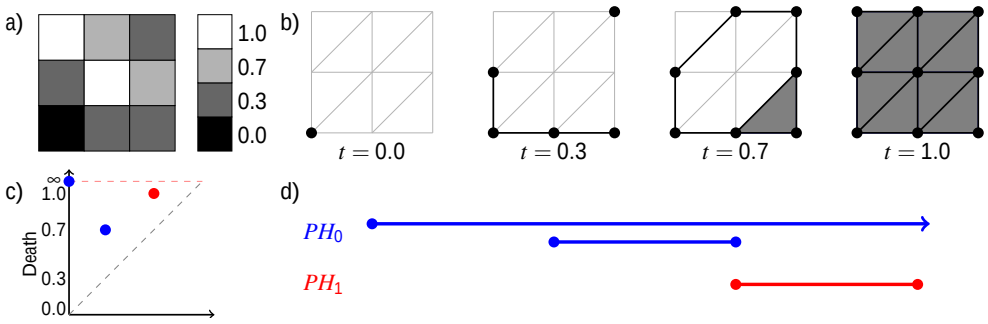

Figure 1: a) Example 3×3 patch from an image. b) Sub-level set filtration at different pixel values t. c) Persistence diagram of the filtration. d) Persistence barcode of the filtration. Persistence pairs are $PH_0 : \{(0, \infty), (0.3, 0.7)\}$ and $PH_1 : \{(0.7, 1.0)\}$. The longer pair in PH_0 represents the connected component that is present throughout, and the shorter pair represents the component that appears at 0.3 and merges with the first component at 0.7. In PH_1, the pair represents the cycle that appears at 0.7. **[NR]**

the value at which the hole appears, and d_i is the value at which the hole is completely filled in. These sets of birth death-pairs are often visualized using persistence barcodes (Figure 1d), in which each pair (b_i, d_i) is represented as a bar that begins at b_i and terminates at d_i, or using persistence diagrams (Figure 1c), in which each pair is plotted as a point on the plane. One of the appealing qualities of persistent homology for texture classification is that birth-death pairs do not change much when the image is deformed, such as by perturbation of pixel values (Cohen-Steiner et al., 2007), and also rotations and scaling of the image.

Once the birth-death pairs for a filtration have been computed, we seek to turn them into features that can be fed into standard machine learning algorithms. Several approaches are suggested in the literature (Adcock et al., 2016; Bubenik, 2015; Adams et al., 2017), and we elect to use a straightforward implementation of Adcock et al. (2016). Fix the homology dimension (either 0 or 1) and consider the set of birth death pairs $\{(b_i, d_i)\}_{i \in J}$ in that dimension. We then compute several polynomials of the form

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

for different choices of α. Polynomials of this form are desirable because they do not depend on the ordering of the pairs, and Adcock et al. (2016) show that they can distinguish spaces with different persistent homology in a meaningful way.

METHOD AND RESULTS

We will describe our method using the publicly available LANDMASS (Alaudah et al., 2015) datasets which consists of two sets of seismic image patches, LANDMASS-1 and

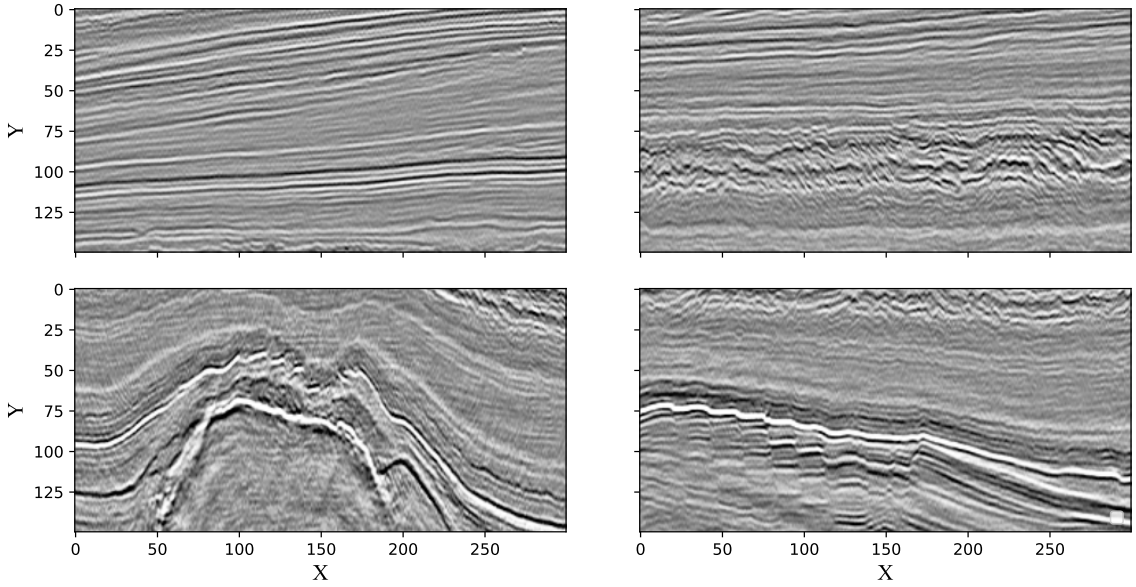

Figure 2: Sample image patches from the LANDMASS-2 dataset in clockwise order from top-left belonging to classes 1, 2, 3, and 4 respectively. **[ER]**

LANDMASS-2, each of which is divided into four classes — horizon patches, chaotic patches, fault patches, and salt dome patches. We will refer to these classes as Class 1, Class 2, Class 3, and Class 4 respectively (see Figure 2 for example images). The image pixel values are in the range [-1,1].

Our method is motivated by the observation that the persistence diagrams of images belonging to different classes exhibit subtle differences in the distribution of persistence pairs, as seen in Figure 3, and hence the polynomial features extracted from them should be able to detect these differences. We start by taking a raw grayscale image, and compute its persistent homology using a sub-level set filtration. The pair in PH_0 with infinite death time is discarded, and 15 polynomial features are calculated per homology dimension, with $\alpha_{j,k} = \delta_{j=j_0,k=k_0}$, where $(j_0, k_0) \in \{0,1,2,3\}^2 - \{(0,0)\}$. This produces a total of 30 topological features which are then fed into one of three classification algorithms — a multi-class support vector machine (SVM), a random forest (RF), and a fully connected neural network (NN) classifier. The NN architecture contains 1 hidden layer with 100 nodes, and an output layer with 4 nodes, one for each class.

The classifiers are trained using 70% of the data from each class chosen randomly, and the remaining 30% are used as the test set to calculate the classification accuracy. We used 5000 epochs to train the NN with the Adam optimizer. For LANDMASS-1 we achieved an overall accuracy of 74.92%, 98.49%, and 99.68% for the SVM, RF, and NN classifiers respectively, while for LANDMASS-2 the corresponding numbers are 79.42%, 98.58%, and 98.5% (also see Figure 5). The SVM classifier performs quite poorly for both datasets, suggesting the need for nonlinear decision boundaries. The

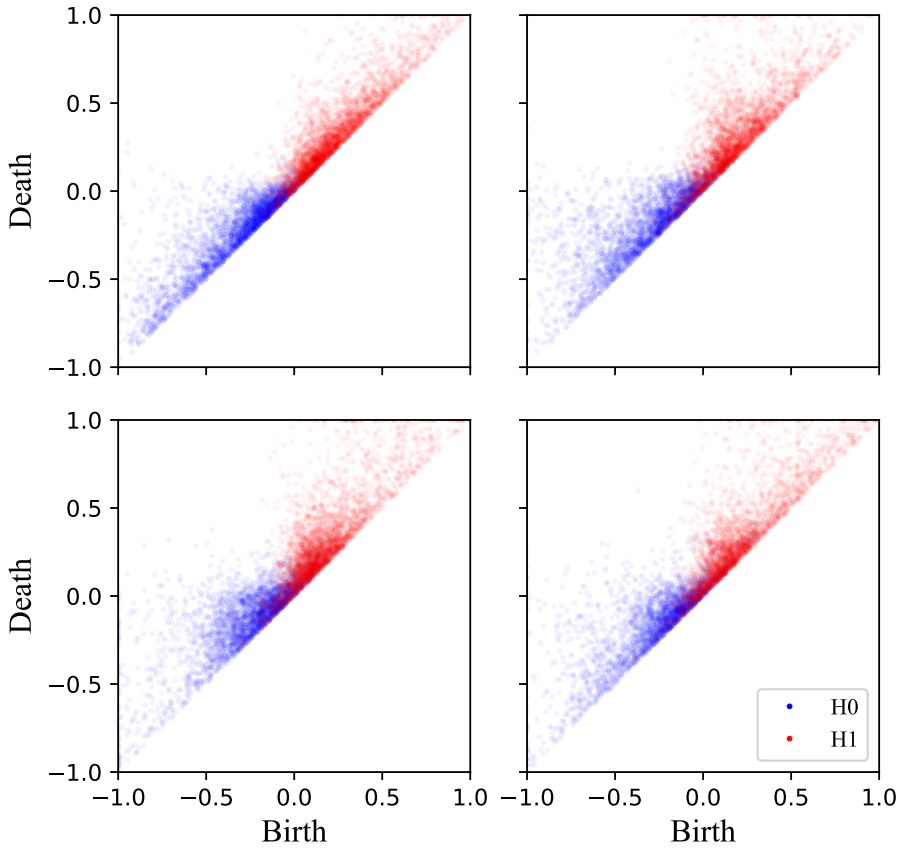

Figure 3: The persistence diagrams in clockwise order from top-left belonging to classes 1, 2, 3, and 4 respectively. The corresponding images are shown in Figure 2. **[CR]**

RF and NN classifiers which are capable of handling non-linear decision boundaries work very well and have similar performance. In Figure 4 we have provided scatter plots of the first two principal components of the topological features for the two datasets, which already begin to show separation of the 4 classes.

We have also explored whether any of the specialized attributes widely used in texture based segmentation of seismic images has any impact on our proposed workflow, when used in place of the raw image. The attributes chosen for this purpose are root mean square amplitude (RMS), and 10 extremely popular attributes based on grey level co-occurrence matrices (GLCM) (Eichkitz et al., 2013; Di et al., 2017). For each of these attributes, we take the image and generate the corresponding attribute images of the same size, and then repeat our workflow with the 3 classifiers. In this abstract, we only report the classification accuracy for the best 4 performing attributes with respect to the RF classifier, along with the raw image for comparison in Figure 5. We see that the performance of the different attributes are similar to each other and the raw image, but the raw image almost always slightly outperforms the attributes. Also very interestingly, for the SVM classifier, GLCM Variance far

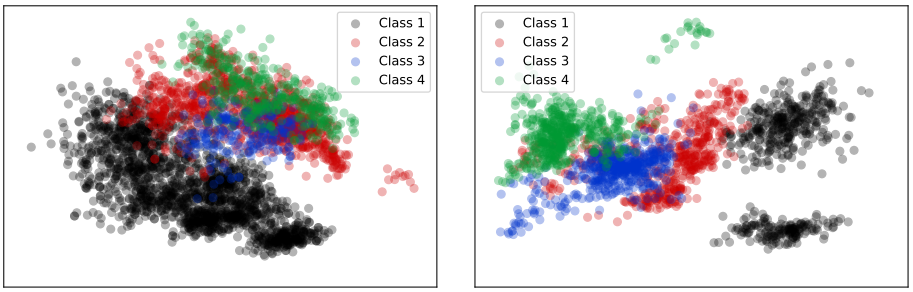

Figure 4: Top two principal components of the topological features for LANDMASS-1 (left) and LANDMASS-2 (right). Separation is more apparent in LANDMASS-2 due to balanced class sizes. **[CR]**

Attribute	Classification Accuracy on Test Set (%)		
	SVM	RF	NN
Raw	99.8 / 75.2 / 0.0 / 0.0	99.9 / 98.6 / 95.2 / 93.3	100.0 / 99.6 / 99.7 / 98.4
Image	100.0 / 55.0 / 88.3 / 74.3	100.0 / 98.0 / 100.0 / 96.3	100.0 / 100.0 / 99.0 / 95.0
GLCM	100.0 / 18.6 / 34.1 / 29.3	99.9 / 97.9 / 82.1 / 93.3	100.0 / 97.8 / 92.8 / 97.0
Mean	62.7 / 19.0 / 4.0 / 100.0	100.0 / 97.0 / 97.3 / 91.7	100.0 / 96.0 / 95.7 / 96.3
RMS	100.0 / 1.0 / 0.0 / 0.0	99.3 / 96.1 / 88.0 / 82.0	99.5 / 99.1 / 96.3 / 91.5
Amplitude	74.7 / 85.7 / 71.3 / 61.7	99.7 / 96.0 / 96.0 / 91.7	99.7 / 99.0 / 93.7 / 91.3
GLCM	100.0 / 0.0 / 0.0 / 0.0	99.3 / 94.9 / 80.8 / 91.2	99.8 / 93.6 / 87.7 / 96.7
Correlation	64.7 / 32.0 / 89.3 / 32.3	99.7 / 93.7 / 92.0 / 97.0	100.0 / 95.7 / 93.7 / 98.3
GLCM	96.6 / 94.1 / 92.8 / 67.7	98.5 / 95.7 / 96.3 / 74.0	99.3 / 98.3 / 98.1 / 87.3
Variance	97.3 / 93.3 / 91.7 / 87.0	99.0 / 95.3 / 96.7 / 89.7	99.7 / 99.0 / 99.3 / 95.0

Figure 5: Classification accuracy for the best 4 texture attributes (with respect to the RF classifier), and for each classification algorithm are given for comparison along with the raw image. The results are broken down for each of the 4 classes, color coded as Class 1 / Class 2 / Class 3 / Class 4. In each cell, the top and bottom rows correspond to the LANDMASS-1 and LANDMASS-2 datasets respectively. **[NR]**

outperforms all other attributes.

CONCLUSIONS

We have presented a new approach for seismic texture classification using features generated from topological data analysis. Experiments on the LANDMASS datasets in conjunction with black box ML algorithms suggest that these features themselves are excellent descriptors of textural information in seismic images. The topological features can be used for detection and classification of important geologic structures of interest, such as faults and salt domes. We anticipate that their inclusion in existing ML workflows for similar tasks will greatly enhance their performance. Future work will aim at extending the proposed method to 3D labeled datasets.

ACKNOWLEDGEMENTS

All persistent homology calculations in this work were implemented using the GUDHI software library (Maria, 2015). The random forest and SVM classifiers were implemented using scikit-learn (Pedregosa et al., 2011), while the neural network classifier was implemented using Tensorflow (Abadi et al., 2015). R.S. was partially supported by the Stanford Exploration Project. B.N. was partially supported by the US DoD NDSEG fellowship program. Both authors would like to thank Gunnar Carlsson for valuable discussions, and Biondo Biondi for suggesting the dataset.

REFERENCES

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, 2015, TensorFlow: Large-scale machine learning on heterogeneous systems. (Software available from tensorflow.org).

Adams, H., T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta, and L. Ziegelmeier, 2017, Persistence images: A stable vector representation of persistent homology: Journal of Machine Learning Research, **18**, 1–35.

Adcock, A., E. Carlsson, and G. Carlsson, 2016, The ring of algebraic functions on persistence bar codes: Homology, Homotopy and Applications, **18**, 381–402.

Alaudah, Y., Z. Wang, Z. Long, and G. AlRegib, 2015, `http://cegp.ece.gatech.edu/codedata/LANDMASS/index.html`.

Bubenik, P., 2015, Statistical topological data analysis using persistence landscapes: Journal of Machine Learning Research, **16**, 77–102.

Carlsson, G., 2009, Topology and data: Bull. Amer. Math. Soc. (N.S.), **46**, 255–308.

Chen, Q., and S. Sidney, 1997, Seismic attribute technology for reservoir forecasting and monitoring: The Leading Edge, **16**, 445–448.

Chevitarese, D. S., D. Szwarcman, E. V. Brazil, and B. Zadrozny, 2018, Efficient classification of seismic textures: 2018 International Joint Conference on Neural Networks (IJCNN), IEEE, 1–8.

Chopra, S., and V. Alexeev, 2006, Applications of texture attribute analysis to 3d seismic data: The Leading Edge, **25**, 934–940.

Chopra, S., and K. J. Marfurt, 2005, Seismic attributes – a historical perspective: Geophysics, **70**, 3SO–28SO.

Cohen-Steiner, D., H. Edelsbrunner, and J. Harer, 2007, Stability of persistence diagrams: Discrete & Computational Geometry, **37**, 103–120.

Di, H., M. Shafiq, and G. AlRegib, 2017, Multi-attribute k-means cluster analysis for salt boundary detection: Presented at the 79th EAGE Conference and Exhibition 2017.

Eichkitz, C. G., J. Amtmann, and M. G. Schreilechner, 2013, Calculation of grey level co-occurrence matrix-based seismic attributes in three dimensions: Computers & Geosciences, **60**, 176–183.

Freudenthal, H., 1942, Simplizialzerlegungen von beschrankter flachheit: Annals of Mathematics, **43**, 580–582.

Ghrist, R., 2017, Homological algebra and data, *in* The Mathematics of Data: volume **25** *of* IAS/Park City Mathematics Series, 273–325.

Love, P., and M. Simaan, 1984, Segmentation of stacked seismic data by the classification of image texture, *in* SEG Technical Program Expanded Abstracts 1984: Society of Exploration Geophysicists, 480–482.

Maria, C., 2015, Filtered complexes, *in* GUDHI User and Reference Manual: GUDHI Editorial Board.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, 2011, Scikit-learn: Machine learning in Python: Journal of Machine Learning Research, **12**, 2825–2830.

Waldeland, A. U., A. C. Jensen, L.-J. Gelius, and A. H. S. Solberg, 2018, Convolutional neural networks for automated seismic interpretation: The Leading Edge, **37**, 529–537.

Zhao, T., V. Jayaram, A. Roy, and K. J. Marfurt, 2015, A comparison of classification techniques for seismic facies recognition: Interpretation, **3**, SAE29–SAE58.