CATSEYES: Categorizing Seismic structures with tessellated scattering wavelet networks

Yash Bhalgat[1,2], Jean Charléty[1], Laurent Duval[1]

[1]IFP Energies nouvelles
92852 Rueil-Malmaison, France
{firstname.lastname}@ifpen.fr

[2]Department of Electrical Engineering - IIT Bombay
Mumbai, Maharashtra 400076, India
yashsb@umich.edu (now with University of Michigan)

http://www.laurent-duval.eu/opus-cats-eyes-seismic-data-classification-scattering-networks.html

Objectives

- Automating the labeling of *structural monads*
- Guarantee an invariance to gradual local deformations
- Ensure sparse representation to ease classification
- Propose a comparison of dimension reduction methodologies

Introduction

To produce sound images of underground structures, geophysics acquire, model and process huge sets of seismic traces, ending up in stacked or migrated datasets (Fig. 1). The latter represent (distorted because indirect) geological formations in the shape of various seismic patterns within the wiggling bandpass nature of seismic signals. Their analysis is of primary importance to understand the tectonic and sedimentary history of regions, and their potential in finding hydrocarbon traps.

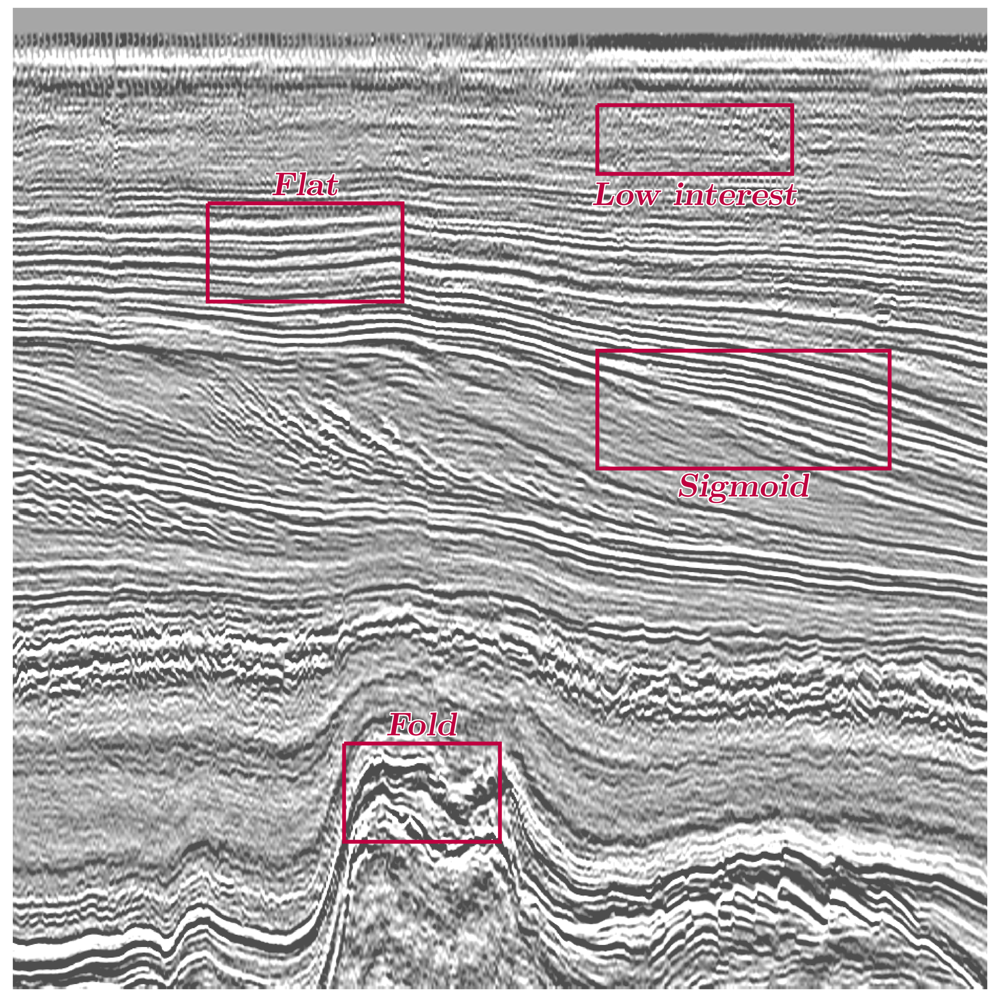


Figure 1 - **Migrated seismic sections, with exemplars of four instances of structural monads: *Flat*, *Sigmoid*, *Fold*, *Low interest*.**

Seismic database evaluation methodology

1. Cropping
2. Curation
3. Repetition removal

The database contains 580 exemplar images of 512×512 pixels divided in 4 classes as:

Classes	Number of images
Flat	221
Fold	223
Sigmoid	100
Low interest	36
Total	580

Methods

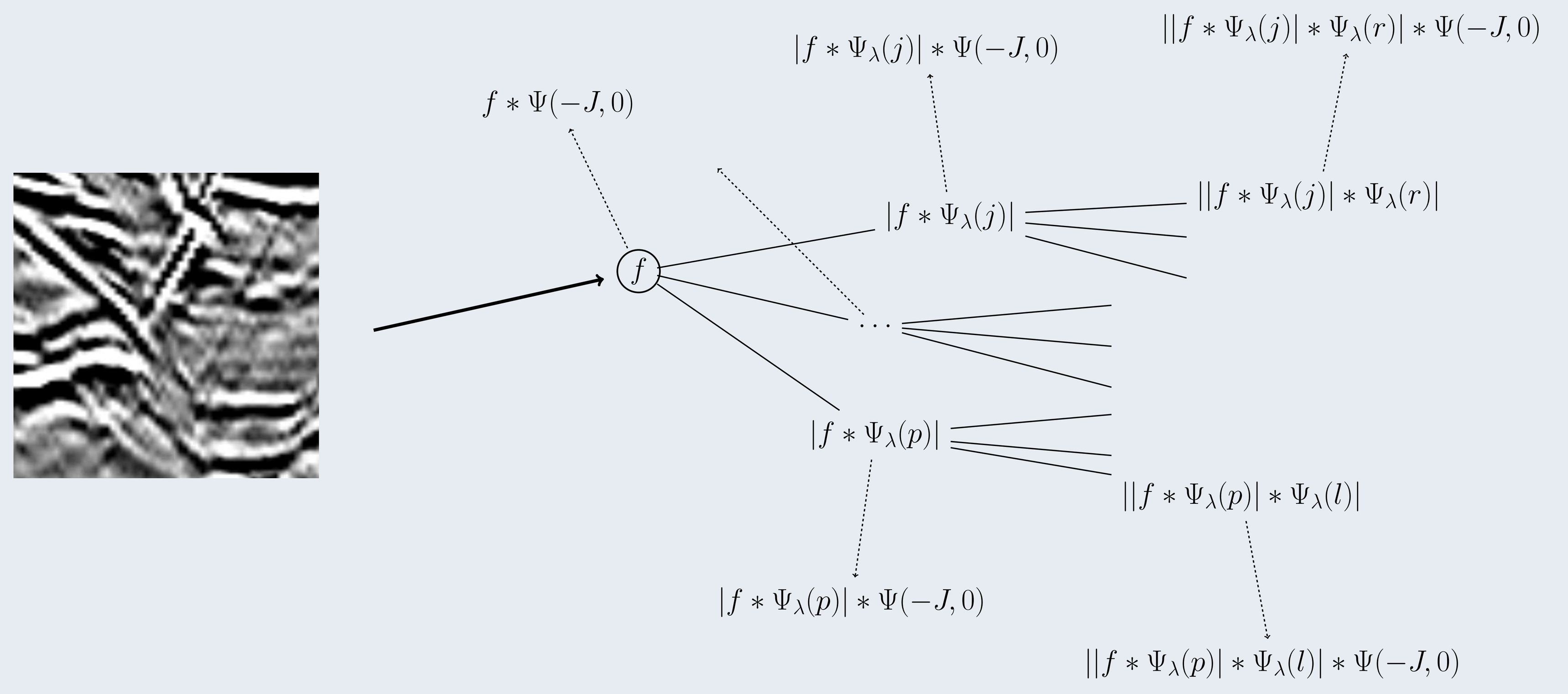
The most natural manner to construct ***feature vectors (FVs)*** — to feed classifiers — reduces to using ***all scattering coefficients***. As they result from many couples of orientations and scales, their size can become extremely large. Exemplar FVs depicted in Figures 3(d-h-l) have a total length of about 1.5×10^7 coefficients.
However, we observe that the FVs are highly compressible. As ScatNets are ***energy preserving***, the steep decay indicates that ***most of the information is carried by very few important coefficients***. Slightly differing decay regimes can be observed across the different classes.

Tessellated scattering networks

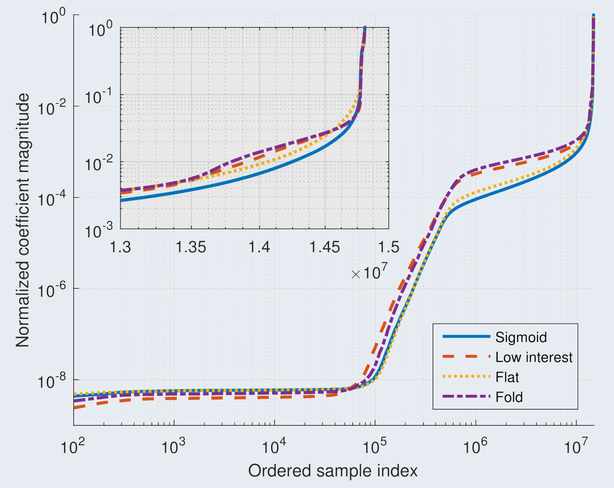
The *beauty* of seismic structures is that the Sigmoid class can be viewed as a combination of two classes: Fold and Flat and a Fold or Sigmoid behavior may slowly warp to a Flat morphology. Hence, we extract and combine FVs from a tessellation of subparts of initial images. This operation can be thought as a diversity enhancement to account for gradual morphing between structural monads. In this work,

1. Images (of size 512×512) are divided into $4 \times 4 = 16$ non-overlapping blocks.
2. Scattering wavelet transform coefficients are extracted from each block.
3. Computing the mean of each of the convolutions was shown [2, 3] to correspond to the energy of the convolutions.

Feature extraction



How **sparse** can the **scattering transform** be ?



Compressibility of ordered scattering transform feature vectors in log-log-scale.

Figure 2 - Scattering network

Scattering transform

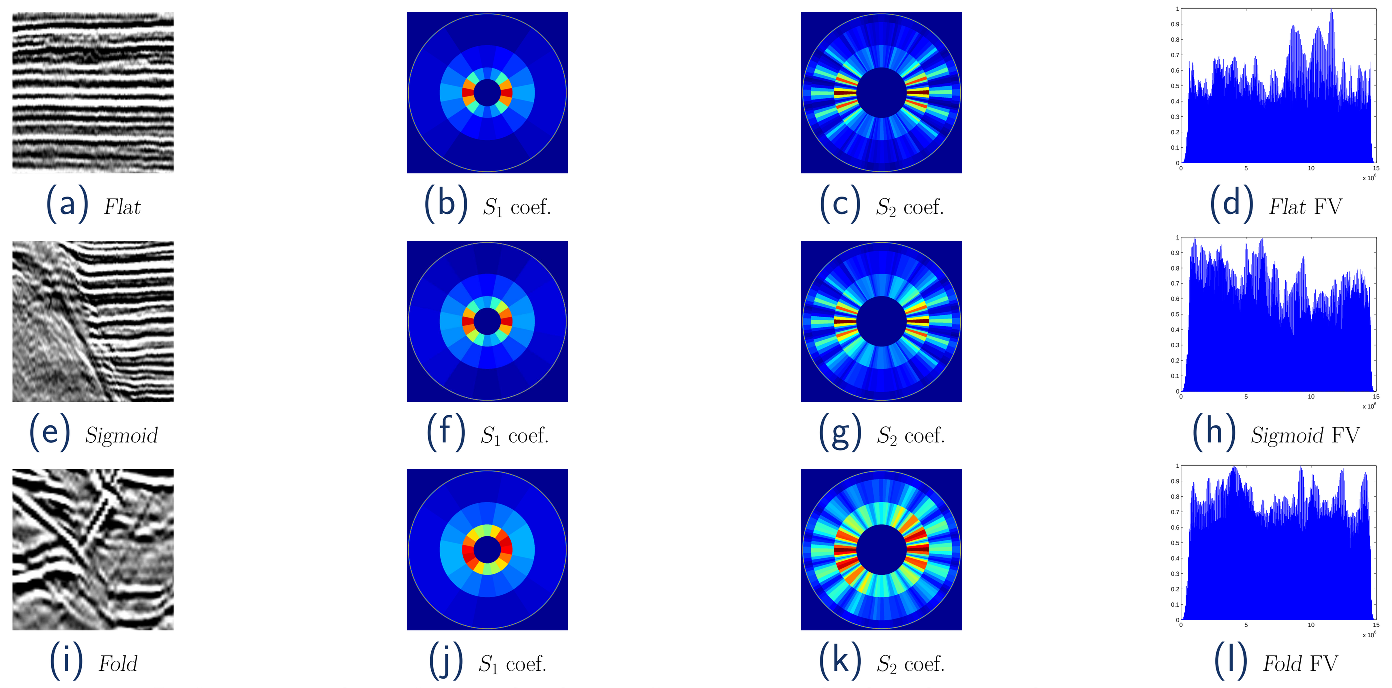


Figure 3 - Left: exemplars for each class. Center: multi-level, angular sector representation of a two-level scattering transform. Right: flattened feature vectors.

Results

Image input 512×512, J=3, 8 angles, final descriptor size is $1.5\ 10^7$.

Comparison of sparsification methods

Table 1 - Accuracy/computational results and comparisons for different dimension reduction and feature extraction methods for varying training percentages.

Training		30%	50%	70%	Time (s) for 50% training		
Method	#RFS	Accuracy (%)			Feat. extract.	Train.	Classif.
Gini	8725	62.4	64.9	65.2	7113	968	0.13
χ^2	7167	57.0	61.2	61.7	5623	797	0.11
CFS	5133	69.0	69.5	70.4	4784	557	0.17
KW	3133	69.1	71.6	71.8	2126	289	**0.10**
mRMR	4607	75.2	76.4	77.9	5275	782	0.31
SBMLR	3265	73.8	75.1	76.1	8982	1044	0.22
Fisher	2819	80.7	81.3	82.5	1931	376	0.87
ScatNet	216	86.6	87.1	87.6	**1814**	**54**	0.47
Tessellated	3456	**90.9**	**91.6**	**93.5**	2214	341	0.14

To exploit redundancy in these highly sparse vectors, we first benchmark various feature selection methods [4] collected at Arizona State University[a]. The feature selection methods applied to our database are based on: Gini index [5], χ^2 statistics [6], Correlation-based Feature Selection (CFS) [7], Kruskal-Wallis (KW) [8], Minimum Redundancy and Maximum Relevance (mRMR) [9], sparse multinomial logistic regression algorithm with Bayesian regularisation (SBMLR) [10] and Fisher score [11]. Reduced feature sizes (#RFS) are globally shrunk again by an order of magnitude, toward thousands of coefficients

Results for the best sparsification method

Table 2 - Confusion matrix for 50 % training. Horizontal: true class; vertical: assigned class.

	Flat	*Fold*	*Sigmoid*	*Low int.*
Flat	103	6	2	0
Fold	5	102	3	2
Sigmoid	1	3	14	0
Low int.	1	2	0	47

Conclusion

- We adopt scattering wavelet networks as deformation and translation invariant joint feature extractors and classifiers.
- A database with tagged *structural monads* is devised, drawn on public data, which could be shared for other publishable studies.
- An extensive comparison of feature vector sdimension reduction methods for classification is performed
- The proposed tessellated scattering decomposition is shown to be effective.

References

[1] P. Birajadar, Y. Bhalgat, U. Sharma, M. Haria, B. Singh, and V. Gadre, “A scattering wavelet network based approach to fingerprint classification,” *Patt. Rec. Lett. (in review)*, Oct. 2017.

[2] S. Mallat, “Group invariant scattering,” *Comm. Pure Appl. Math.*, vol. 65, no. 10, pp. 1331–1398, Oct. 2012.

[3] L. Sifre and S. Mallat, “Rotation, scaling and deformation invariant scattering for texture discrimination,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, Portland, OR, USA, 2013.

[4] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, and H. Liu, “Advancing feature selection research — ASU feature selection repository,” Tech. Rep., 2010.

[5] C. Gini, *Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche*, 1912.

[6] H. Liu and R. Setiono, “Chi2: feature selection and discretization of numeric attributes,” in *Proc. 7th IEEE Int. Conf. Tools with Artificial Intelligence*, Herndon, VA, USA, Nov. 1995.

[7] M. A. Hall and L. A. Smith, “Feature selection for machine learning: Comparing a correlation-based filter approach to the wrapper,” in *Proc. Int. Florida Artificial Intelligence Research Soc. Conf.*, Orlando, FL, USA, May 1-5, 1999.

[8] L. J. Wei, “Asymptotic conservativeness and efficiency of Kruskal-Wallis test for K dependent samples,” *J. Am. Stat. Assoc.*, vol. 76, no. 376, pp. 1006–1009, Dec. 1981.

[9] H. Peng, F. Long, and C. Ding, “Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[10] G. C. Cawley, N. L. C. Talbot, and M. Girolami, “Sparse multinomial logistic regression via Bayesian L1 regularisation,” in *Proc. Ann. Conf. Neur. Inform. Proc. Syst.*, Dec. 4-7, 2006, pp. 209–216.

[11] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, 2nd edition, 2001.

[a] featureselection.asu.edu/old/software.php