A CURVELET-BASED DISTANCE MEASURE FOR SEISMIC IMAGES

Yazeed Alaudah and Ghassan AlRegib

Center for Energy and Geo Processing (CeGP) at Georgia Tech and KFUPM
School of Electrical and Computer Engineering
Georgia Institute of Technology
{alaudah,alregib}@gatech.edu

ABSTRACT

We introduce a new curvelet-based distance measure for post-migration seismic data. The measure exploits the highly directional content of seismic images. It calculates the sum of the squared chord distances between the histograms of the curvelet coefficients of two images, over all orientations and scales. In a retrieval task consisting of 918 retrieval instances, the proposed method successfully retrieves all images. This wasn't achieved by other methods in the literature. Furthermore, in comparison to the state-of-the-art, the proposed measure required 86% less computation time.

1. INTRODUCTION

Modern seismic surveys produce extremely big data. There are great challenges in terms of processing, classifying, and organizing these big data. Efficient and robust distance measures are required to enable the quantification of the difference between two datasets or images in a way that is meaningful. One application of these seismic distance measures is that they make it possible to retrieve datasets that contain similar seismic structures, leading to increased efficiency in storage, compression, and processing. It can also be particularly useful for clustering big seismic data, where typical distance measures such as the mean square error (MSE) don't describe the data very well.

The focus of this paper is on post-migration seismic images specifically. These are gray scale images that result from the stacking and migration of seismic traces obtained through a regular seismic survey. These images are used to visualize and analyze subsurface geological structures in order to assess the prospects of oil- and gas-bearing reservoirs. An example of these images is shown in Fig. 3.

Several authors have proposed distance, or similarity, measures for seismic data. Benvegna *et al.* [1] introduced a seismic dissimilarity measure for seismic traces based on the cross-correlation of the two signals, as well as their cumulative shape dissimilarity. However, it was applied on seismic traces, and not migrated seismic images. Al-Marzouqi *et al* [2] introduced a similarity index for seismic images using adaptive curvelets. The scale and angle parameters are found by an optimization algorithm that maximizes the coefficient of variation of the curvelet coefficients [3]. Long *et al.* [4] introduced a seismic similarity measure, SeiSIM, that uses a texture similarity metric (STSIM-1) that was introduced by [5] combined with a geological attribute obtained from discontinuity maps.

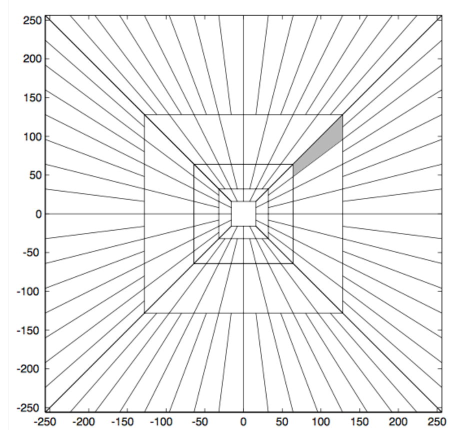


Fig. 1: Curvelet tiling of the frequency spectrum showing different scales and orientations; adapted from [6].

In this paper, we propose a distance measure for seismic images that is based on curvelets. We show that this measure performs as good as the state-of-the-art, while only using 86% less computation time. The rest of the paper is organized as follows: An overview of the curvelet transform is presented in section 2. This is followed by the technical details of our proposed distance measure in section 3, as well as experimental results in section 4.

2. THE CURVELET TRANSFORM

The curvelet transform is a directional multiscale decomposition that was first introduced by Candés *et al.* [6]. It provides an efficient way to represent images with high directional content. It has been shown [7] that images that contain geometrically regular edges are more compactly represented by curvelet frames rather than wavelet bases. This is especially true for seismic data, where the wavefronts lie mainly along smooth curves.

The curvelet transform works by taking the 2D fast Fourier transform of an image (2D FFT) and then dividing the plane into multiple scales and orientations as is shown in Fig. 1. The total number of scales in the curvelet tiling, J, depends on the size on the image, and is given by:

$$J = \lceil \log_2 \min(N_1, N_2) - 3 \rceil, \tag{1}$$

where N_1 and N_2 are the number of pixels in vertical and horizontal directions, respectively; and $\lceil * \rceil$ is the ceiling function. The number of orientations at scale $j \geqslant 1$, $K(j)$, is given by:

$$K(j) = 16 \times 2^{\lceil (j-1)/2 \rceil}. \tag{2}$$

For scale $j = 0$, there is only one orientation. Curvelets obey a parabolic scaling relationship that insures that the dimensions of the support of a curvelet element follows $width \approx length^2$ [7].

Curvelet coefficients are then generated by taking the IFFT for each wedge (such as the one highlighted in figure 1) after multiplying it by a smoothing function. Since the FFT of real images is symmetric around the origin, only two quadrants of the Fourier spectrum are necessary for obtaining the curvelet coefficients.

3. PROPOSED DISTANCE MEASURE

To obtain a reasonable distance measure for seismic images, we must ensure that the emphasis of the measure is on the seismic substructures, and not the actual pixel values of the images. As mentioned in section 2, curvelets offer a compact representation of seismic data, where most of the energy of the data is concentrated in a few coefficients. Our general approach is to take the squared chord distance [8] between corresponding histograms of curvelet coefficients, at all scales and orientations. An outline of the proposed algorithm is shown in Fig. 2.

The first step, normalization, involves normalizing the pixel values of the two images by subtracting the mean, and dividing by the standard deviation. This insures that the distance measure is invariant to the absolute values of the pixels, and rather to underlying structure of the two images. This can be expressed as:

$$\tilde{f} = \frac{f - \mu_f}{\sigma_f}, \tag{3}$$

where f is the original image, $\tilde{f}$ is the normalized image, μ_f and σ_f are the mean and standard deviation of f, respectively.

After applying the forward curvelet transform to $\tilde{f}$, as explained in section 2, the squared-chord distance [8] is used to calculate the distance between the corresponding histograms of the two images for each orientation and scale. Let $H^f_{j,k}(i)$ be the i^{th} bin in the histogram of the curvelet coefficients of image f at scale j and orientation k. Similarly, $H^g_{j,k}(i)$ for image g. The squared-chord distance between the two histograms is then defined as:

$$d_{SC}(H^f_{j,k}, H^g_{j,k}) = \sum_{i=1}^{M} \left(\sqrt{H^f_{j,k}(i)} - \sqrt{H^g_{j,k}(i)} \right)^2, \tag{4}$$

where M is the total number of bins in the histogram. The final distance measure between the two images can then be expressed as:

$$D(f,g) = \sum_{j=0}^{J} \sum_{k=1}^{K(j)/2} w(j) \cdot d_{SC}(H^f_{j,k}, H^g_{j,k}), \tag{5}$$

where J is the number of scales defined in Eq. 1. $K(j)$ is the number of orientations at scale j, and $w(j)$ is a weighing vector for different scales. The default values for $w(j)$ are all ones except zero at the highest scale, since the highest scale usually corresponds to small details, or noise, that do not affect the overall structure of the images. The measure takes the squared chord distance over the *corresponding* histograms in each scale and orientation, and thus is not scale or rotation invariant. This allows the measure to be sensitive to structures of different orientations and scales, such as small discontinuities and large faults.

4. EXPERIMENTAL RESULTS

We evaluate the performance of our distance measure by using a simple image retrieval test, and compare its performance to well known image distance or similarity measures such as Mean Square Error (MSE) and SSIM [9], as well as SeiSIM [4] and the method introduced by Al-Marzouqi *et al* [2].

4.1. Data

We use a dataset of $N = 54$ seismic images of size 100×200 that have been extracted from the Netherlands offshore F3 database (available online [10]). The images were hand labeled into 3 classes. Each class containing $K = 18$ different images. These classes are: `Horizon`, `Fault`, and `Salt Dome`. An example from each of class is shown in Fig. 3. Each one of the images was extracted from a different inline section. This is to ensure that while images have the same structure, they aren't similar on the pixel level.

4.2. Performance Evaluation

For each image f_i in our database that belongs to class $c_i \in \{\texttt{Horizon}, \texttt{Fault}, \texttt{Salt Dome}\}$, we retrieve 17 images, g_k for $k = 1, 2, \cdots, K-1$ from the pool of N images where

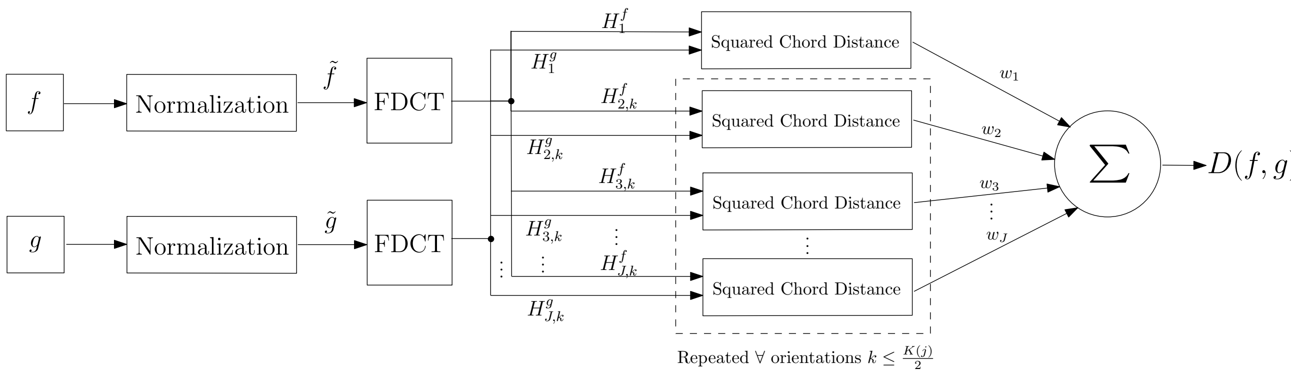
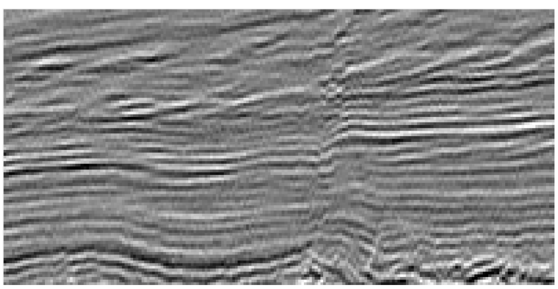
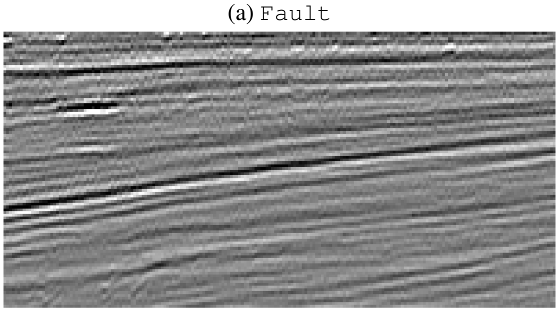


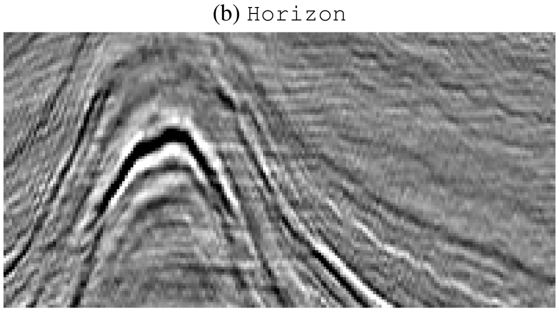
Fig. 2: Block diagram of the proposed method, FDCT refers to the forward discrete curvelet transform.



(a) `Fault`



(b) `Horizon`



(c) `Salt Dome`

Fig. 3: Examples of the three different classes of images in the database

K is the number of images per class, and N is the total number of images. The classes of the query and retrieved images are assumed to be unknown during the retrieval. Retrieval is done by ranking the images based of their $D(f_i, g_k)$ values in an ascending order, and choosing the first $K-1$ images as the *relevant* images. We then label these g_k's as having the same class as f_i, i.e. c_i. We repeat this for all images in the database, and then find the *retrieval accuracy* (RA) defined as:

$$\text{RA} = \frac{100 \times \sum_{i=1}^{N} \sum_{k=1}^{K-1} \mathbf{1}_{\{f_i, g_k \in c_i\}}}{N \times (K-1)}. \tag{6}$$

In other words, the retrieval accuracy is just the ratio of the number of correctly retrieved images to the total number of images that were retrieved. There are 918 (i.e. $N \times (K-1)$) instances where images are being retrieved based on their distance to a reference image f_i in the database.

We also calculate the values of the following measures that are typically used for the evaluation of image retrieval measures [5]. These measures are all valued from 0 to 1, with larger values corresponding to better retrieval performance.

- Precision at One (P@1): is the number of times that the first retrieved image is of the same class as the query image divided by the total number of retrieved images (i.e. $K-1$). Precision at One is typically used when one is interested in the first retrieved image only.
- Mean Average Precision (MAP): the average precision (AP) scores for each query is given by:

$$\text{AP} = \frac{\sum_{\forall k} P(k) \times Rel(k)}{N_r} \tag{7}$$

where $Rel(k)$ is a binary function that equals 1 if the retrieved image at rank k is a relevant image (i.e. of the same class) and zero otherwise. $P(k)$ is the precision at rank k, and N_r is the number of relevant images. After AP is calculated for each query image, MAP is found by averaging AP over all query instances.

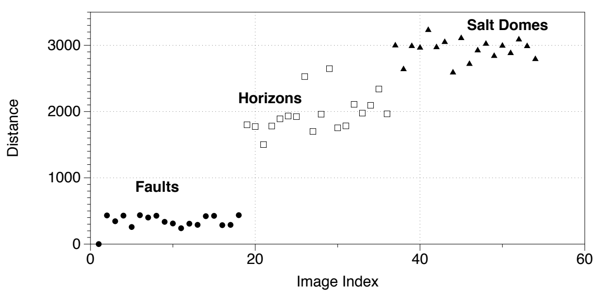


Fig. 4: Distance of all images in the database to a reference `Fault` image, as calculated by the proposed measure

Table 1: Image retrieval results

Metric	P@1	MAP	RA (%)	CPU Time
MSE	0.9630	0.6608	54.14	**1.57**
SSIM	**1**	0.7049	58.93	14.2
Almarzouqi *et al* [2]	0.9630	0.8192	75.27	1758.7
SeiSIM [4]	**1**	0.9949	98.69	1451.7
Proposed Method	**1**	**1**	**100**	197.39

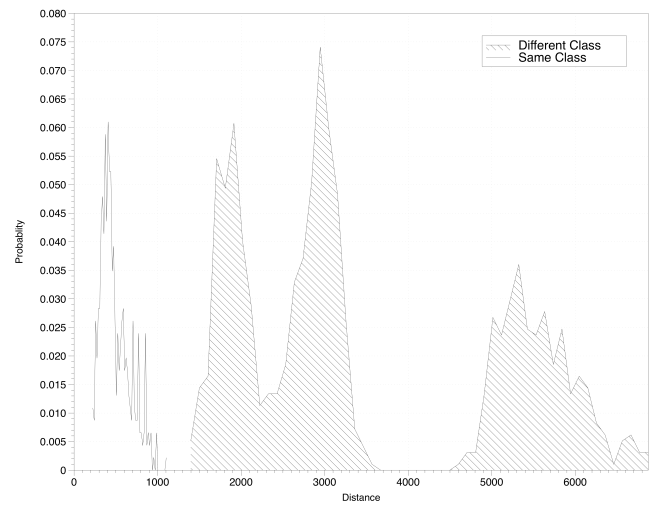


Fig. 5: Probability Density Functions of the distances (calculated using the proposed method) between query and retrieved images when they are of the same (white) or different (shaded) class.

MAP and RA are more suitable choices to measure the retrieval performance when there are more than one relevant retrieved image, as is the case here. The difference is that MAP takes the order of the retrieved images into consideration, while RA doesn't.

4.3. Results

We ran the experiments using the default values for the $w(j)$. We compare the distance measure we proposed to typical image distance or similarity measures such as MSE and SSIM, as well as two other seismic similarity measures, SeiSIM [4] and the method introduced by Al-Marzouqi *et al* [2]. The results are shown in table 1.

It's clear from table 1 that SeiSIM and our method significantly outperform the other measures. Furthermore, while the values of SeiSIM and our method are very close, our method is more than 7 times faster. This is mainly because our method is computed globally over the whole image, while SeiSIM requires using a moving window for every pixel in the image to calculate the discontinuity map. CPU Time is in seconds and is calculated using MATLAB's `cputime` function, and is used to roughly compare the computation time of the different methods. The CPU Time is the total time to retrieve all 918 images, thus it can be divided by 918 to find the computation time for one retrieval. The calculations were performed on a computer with the following configuration: Intel Core i7 3.4 GHz, with 32GB of RAM, and running on 64-bit windows 7 using MATLAB 2014B.

Furthermore, we show in Fig. 4 the average distances (computed by the proposed method) of all images in the database to a reference `Fault` image. It can be clearly observed that images of the same class are clustered at approximately the same distance from the reference image. This is true for `Horizon` and `Salt Dome` reference images as well. This indicates that the distance measure we proposed discriminates the images based on their relative structural differences, and not their pixel values. This is also apparent in Fig. 5, where we plot the probability distribution function of the distances of each image from other images of the same class (white) and from images that belong to other classes (shaded) over all images in the database. The figure shows a clear distinction between the the two PDFs. By using a simple thresholding operation, we can judge whether two images are of the same class or not based solely on their distance value. This lends itself very well to classification, clustering, and dimensionality reduction applications.

5. CONCLUSION

In this paper we introduced a new distance measure for post-migration seismic data based on the curvelet transform. The distance measure takes the sum of the squared chord distances between the histograms of the curvelet coefficients of two images over all orientations and scales. Experimental results show that this measure is on par with the state of the art seismic similarity measure, but much more computationally efficient, requiring 86% less computation time. For future work, we plan to use this method for clustering different seismic images as well as automated tracking of seismic sub-structures.

6. REFERENCES

[1] F. Benvegna, A. DAlessando, G. Bosco, D. Luzio, L. Pinello, and D. Tegolo, "A new dissimilarity measure for clustering seismic signals," in *Image Analysis and Processing ICIAP 2011*, ser. Lecture Notes in Computer Science, G. Maino and G. Foresti, Eds. Springer Berlin Heidelberg, 2011, vol. 6979, pp. 434–443.

[2] H. Al-Marzouqi, G. AlRegib *et al.*, "Similarity index for seismic data sets using adaptive curvelets," in *2014 SEG Annual Meeting*. Society of Exploration Geophysicists, 2014.

[3] H. Al-Marzouqi and G. AlRegib, "Using the coefficient of variation to improve the sparsity of seismic data," in *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*, Dec 2013, pp. 630–630.

[4] Z. Long, Z. Wang, and G. Alregib, "Seisim: structural similarity evaluation for seismic data retrieval," in *Submitted to IEEE Global Conference on Signal and Information Processing (GlobalSIP) 2015*.

[5] J. Zujovic, T. N. Pappas, and D. L. Neuhoff, "Structural similarity metrics for texture analysis and retrieval," in *Proceedings of the 16th IEEE International Conference on Image Processing*, ser. ICIP'09. IEEE Press, 2009, pp. 2201–2204.

[6] E. Candes, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Modelling and Simulations*, vol. 5, no. 3, pp. 861–899, 2005.

[7] E. J. Candes and D. L. Donoho, "New tight frames of curvelets and optimal representations of objects with piecewise c2 singularities," *Communications on Pure and Applied Mathematics*, vol. 57, no. 2, pp. 219–266, 2004.

[8] S.-H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," *International Journal of Mathematical Models and Methods in Applied Sciences*, vol. 1, pp. 300–307.

[9] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[10] Netherlands offshore f3 block - complete. [Online]. Available: https://opendtect.org/osr/pmwiki.php/Main/NetherlandsOffshoreF3BlockComplete4GB