A SYSTEM FOR SEISMIC DATA PROCESSING

Klaus Köster and Michael Spann
School of Electronic & Electrical Engineering, The University of Birmingham,
Pritchatts Road 52, Edgbaston, Birmingham B15 2TT,UK
Tel: +44 121 414 3685; fax: +44 121 414 4291
e-mail: koesterk@eee.bham.ac.uk, M.Spann@bham.ac.uk

1 INTRODUCTION

This work describes an unsupervised method to extract 2D and 3D inner earth structures from seismic reflection measurements in order to support an expert interpreting the data. The problem is solved with a general texture segmentation approach, which first extracts characteristic patterns for seismic events (feature extraction stage) and then groups data elements with similar patterns to geologic structures (segmentation stage). The result can then be visualised as 3D models to get a better idea of the structures. Experimental results on both, real and synthetic data and conclusions are presented in Section 4 and Section 5.

2 FEATURE EXTRACTION

Taner *et al*[5] discussed the locally emergent frequency as a useful feature to describe seismic events. Here this quantity is estimated via a bank of Gabor filters, since Gabor filters offer the best possible response localisation in the time and frequency domains [1] ensuring a detailed feature map.

2.1 Gabor Filter Banks

The applied filter can mathematically be described as

$$m(x,y,z) = e^{-\frac{1}{2}\left[\frac{x^2}{\sigma^2}+\frac{y^2}{\sigma^2}+\frac{z^2}{\sigma_z^2}\right]} \cos\left(2\pi\mu_0(x\cos\theta + y\sin\theta)\right), \tag{1}$$

where μ_0 denotes the radial centre frequency and θ the orientation of the Gabor wavelet. Equation 1 uses information of the third dimension for smoothing and assumes that frequency parts in this direction are negligible against the contributions in the x- and y-directions. This assumption is valid in data sets, where changes in the z-direction happen to be very smooth. The advantage is the separability of the convolution mask in the z-direction.

For the application to seismic data, a filter bank of 10×10 filters has been designed with frequencies μ_0 between 0 and π and orientations θ between 0 and 180 degrees in the Fourier domain. To estimate the local emergent frequency, the three largest responses of adjacent filters of the filter bank are stored. The local emergent frequency and orientation are computed from the weighted average of these three values for each voxel/pixel in the data set.

3 THE SEGMENTATION ALGORITHM

The segmentation stage in this work uses a region growing technique based on statistical decision rules. This technique segments the entire image into small cells (e.g. 3x3x3 voxels or 5x5 pixels) and merges them iteratively based on a statistical measurement.

3.1 The Data Model

The region data $V = \{\vec{v}_1, \vec{v}_2, \dots\}$ is described by a polynomial description:

$$\vec{v}_i = \vec{a}_0 + \vec{e}_i, \tag{2}$$

where the $\vec{a}_0$ is the model parameter and $\vec{e}_i$ is an independent additive normally distributed noise term. The model parameters are robustly estimated by the *least median of squares* (LMedS) estimator described in [4, 3]. The model's noise term $\vec{e}_i$ is taken into consideration by a covariance matrix $\boldsymbol{C}$ for each region. The smoothness of the estimate is computed by:

$$\text{Smoothness} = \sqrt{Det(\boldsymbol{C})}, \tag{3}$$

where $Det()$ denotes the determinant. To test whether a sample $\vec{x}$ is an inlier within the model, the decision rule

$$\left(\vec{x} - \hat{\vec{v}}_i\right)\boldsymbol{C}^{-1}\left(\vec{x} - \hat{\vec{v}}_i\right)^T \leq \chi^2_{p,0.975} \tag{4}$$

has to be true. $\hat{\vec{v}}$ denotes the LMedS computed estimation of $\vec{v}$. The value $\chi^2_{p,0.975}$ is from chi-square statistics and p denotes the degrees of freedom, which in this case is equal to the dimension of the data vector.

3.2 Initial Stage of the Algorithm

The original data set is tessellated with cells of $3\times3\times3$ voxels for the 3D case and 5×5 pixels for segmentation of 2D images. The LMedS estimator is applied for each cell to estimate the model parameters. This estimator results in more robust results for smaller sets compared to traditional least squares methods [4]. To avoid cells placed on a boundary of two or more regions, the initial cells are shifted by at most one voxel (2 pixels in 2D) per direction to a position resulting in a local minimum of the smoothness criterion (equation 3). The initial cells are overlapping by 2 voxels (4 pixels in 2D) in each dimension to ensure that all voxels (pixels in 2D) are covered by a cell after the shifting procedure. By shifting the cells the tessellation will adapt to any region structure and the final segmentation will not contain the "block-structure" of the cells.

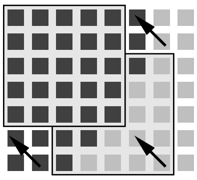

Figure 1: Shifting of the initial cells.

3.3 Merging with the Mutual Inlier Ratio

This region-growing algorithm uses the mutual inlier ratio to derive to a merge decision assuming a Gaussian distribution for the elements of the regions. If R_A is the set of elements of region A, then the set of inlier I_A of region A is

$$I_A = \left\{ \vec{v} \in R_A : (\vec{v} - \vec{v}_A) \boldsymbol{C}_A^{-1} (\vec{v} - \vec{v}_A)^T \leq \chi^2_{p,0.975} \right\}, \tag{5}$$

where $\vec{v}_A$ denotes the estimated model of region A and $\boldsymbol{C}_A^{-1}$ is its inverse covariance matrix. Thus the inlier of A, which are also inlying in B are determined by:

$$I_{AB} = \left\{ \vec{v} \in I_A : (\vec{v} - \vec{v}_B) \boldsymbol{C}_B^{-1} (\vec{v} - \vec{v}_B)^T \leq \chi^2_{p,0.975} \right\}, \tag{6}$$

and the mutual inlier ratio M_{AB} is defined as:

$$M_{AB} = \frac{\mathfrak{card}(I_{AB})}{\mathfrak{card}(I_A)}, \tag{7}$$

where the operator $\mathfrak{card}()$ returns the number of elements of a set. If the computed ratios M_{AB} and M_{BA} are larger than a discriminating value, then the two distributions are found to be similar. Meer [3] uses a specified fixed value as a discriminating value. Smaller sets imply larger associated statistical uncertainties. Therefore the Meer specified value must be a function of the set size, to ensure that the probability of merging two regions is invariant of the set size. The exact estimation of the model parameters suffers from a finite sample size error, which decreases with growing set size n. The relationship between the exact unknown mean of a distribution μ and the estimated mean $\bar{X}$ can be expressed by [2]:

$$\mu = \bar{X} \pm d, \qquad d = t_{0.975} \frac{s}{\sqrt{n}}, \tag{8}$$

where s denotes the estimated standard deviation, n is the set size and $t_{0.975}$ is based on the Student t distribution with a 97.5% confidence interval allowing us to use the estimated standard deviation instead of the true (unknown) standard deviation. Having two estimates of the same distribution, the worst case occurs if estimate $\bar{X}_1$ equals $\mu + d_1$ and estimate $\bar{X}_2$ is computed as $\mu - d_2$. A discrimination value M_{disc} finding the two estimates of the same distribution to be similar can theoretically be determined by solving the integral over the overlapping inlier area of the two estimated distributions. According to the chosen χ^2-value, M_{disc} is 0.975 for a set of infinite size as the finite set error d disappears. It is difficult to find a closed expression for this error approximation, since it still contains the estimated standard deviations s_1 and s_2. Nevertheless, M_{disc} can be determined empirically for different configurations of model degree, spatial dimension of the image and vectorial data dimension. Here two sets of Gaussian distributed independent vectors are generated, their distribution is estimated and the two mutual inlier ratios are stored. This experiment is repeated 400 times to get a reasonable certainty. From the stored inlier ratios the $(1.0 - 0.975) \times 2 \times 400$th *smallest* one is picked as the discrimination value for the current configuration. This measurement has been repeated for set sizes between 20 and 1000. The result is discrimination values for the inlier ratio which allow recognition of identical Gaussian distributed sets with a certainty of 97.5%.

M	S	V	Approximating Function
0	2,3	1	$M_{\text{disc}}(size) = 0.95 - 2.1 \times size^{-0.78}$
0	2,3	2	$M_{\text{disc}}(size) = 0.94 - 26.3 \times size^{-1.37}$
1	2	1	$M_{\text{disc}}(size) = 0.92 - 16.0 \times size^{-1.08}$
1	3	1	$M_{\text{disc}}(size) = 0.92 - 19.9 \times size^{-1.1}$
1	2	2	$M_{\text{disc}}(size) = 0.94 - 81.2 \times size^{-1.58}$
1	3	2	$M_{\text{disc}}(size) = 0.94 - 48.3 \times size^{-1.46}$

Table 1: Approximation functions.

The measured values have been approximated by a function of the form $M_{\text{disc}}(size) = a - b \times size^{-c}$ to avoid storing them explicitly. Table 1 displays approximating functions for configurations of spatial image dimension S, vectorial data dimension V and polynomial model degree M. The parameters a, b and c are computed by least squares regression of the measured values. Figure 2(a) shows the approximating functions for some configurations.

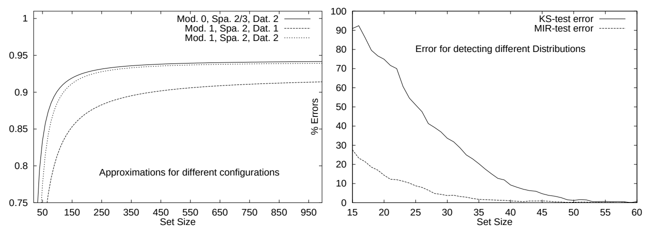

Figure 2: Approximating functions (left) and comparison of mutual inlier ratio with KS-test (right).

To compare the mutual inlier ratio with an established method two Gaussian distributed sets of significantly different mean and variance were created and examined by the KS-test and the mutual inlier test. Figure 2(c) displays the average error rates for different set sizes.

The figure shows a significantly greater reliability when using the mutual inlier ratio for small sets, which is a crucial property for a region-growing algorithm.

3.4 The Iterative Region Merging Stage

This final stage of the algorithm iteratively merges the cells to primitive regions (PR) until no adjacent region pair satisfies the merging rule anymore. Each iterative merging step starts by determining the neighbour resulting in a mutual satisfaction of the decision rule (equation 4) with the *highest* sum of mutual inlier rations MIR:

$$MIR = M_{AB} + M_{BA}. \qquad (9)$$

The stage starts merging the PRs after the best neighbour for merging has been determined for each PR. Each PR satisfying the merge test with one of its neighbours, is linked to this neighbour. The iterative merging stage is implemented using hierarchical graph structure and works in parallel in order to use multiprocessor architectures.

4 RESULTS

Figure 3 displays a 2D synthetic data set and the results obtained from the instantaneous frequency (IF) estimation, a tool applied for seismic trace interpretation, compared with the Gabor bank result. The frequency

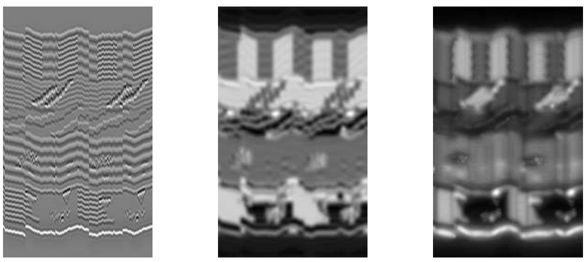

Figure 3: Synthetic seismic image (l.) and results obtained from the IF estimator (m.) and Gabor bank (r.).

measurements are, qualitatively, strongly in agreement with the pattern in the top part of the image. Although the Gabor filter responds with small deviations at the pattern boundaries, details in the bottom part of the image are distinguished slightly better than by the IF approach.

Figure 4 shows a slice of a Gabor filtered real 3D seismic data set and the result obtained from the 3D segmentation stage. The displayed segmentation result contains many detailed regions leaving the final decision to merge regions to the interpreter. In the intended application the financial consequences of over-merging are possibly much worse than for under-merging. Therefore a greater amount of autonomy is allowed for the human interpreter in this case. The displayed 3D model of the two major structures in the bottom half of the image in figure 4 is obtained by manually merging the algorithm's result.

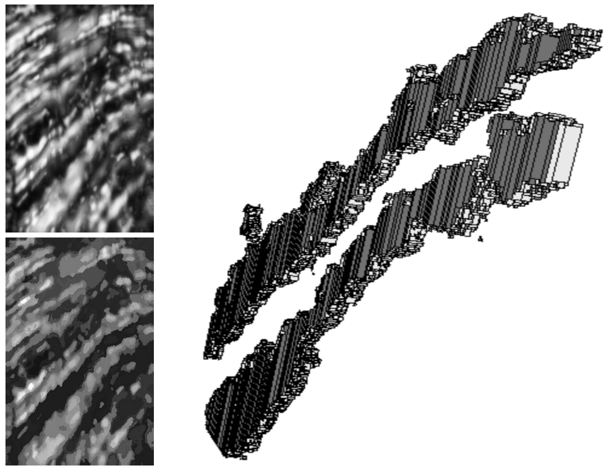

Figure 4: 3D seismic data with result of the Gabor bank, segmentation and a 3D model of the two diagonal regions in the bottom part of the image.

5 CONCLUSIONS

The described system extracts information on geologic structures by applying a bank of Gabor filters, which works with good precision but is computationally demanding since Gabor filter masks are not separable spatially. Nevertheless, as a feature for segmentation it is superior to the IF, which has already proven to be a useful tool for interpreting seismic traces.

The second stage of the system generates a 2D or 3D structure map based on the extracted feature by applying a robust-region growing technique. A new shift technique for the initial cells using the LMedS estimator enables the initial tessellation to adapt to any 2D or 3D region shape. The algorithm does not need any parameters and is completely free from any assumptions on region number or size. A merging technique has been proposed offering a high robustness especially at the crucial initial merging stage of the algorithm, where statistical estimates possess a high uncertainty due to small set sizes.

References

[1] Bovik et al, "Localised Measurement of Emergent Image Frequencies by Gabor Wavelets", IEEE Transact. on Inf. Theory 38, 2, 1992

[2] Lapin, L., "Probability and Statistics for Modern Engineering", 2nd Ed. PWS-Kent, Boston, 1990

[3] Meer et al, "Least Median of Squares based Robust Analysis of Image Structure", Proc. DARPA Image Underst. Workshop, 1990, pp. 231-254,

[4] Rousseeuw et al, "Robust Regression and Outlier Detection", Wiley, New York, 1987

[5] Taner et al, "Complex Seismic Trace Analysis", Geophysics 44, 6, June 1979