IJCAT - International Journal of Computing and Technology, Volume 4, Issue 2, February 2017
ISSN : 2348 - 6090
www.IJCAT.org

A Seismic Fault Recognition Method based on Linear Regression

[1] **Lei CHEN;** [2] **Chuang-Bai XIAO;** [3] **Jing YU;** [4] Zhen-Li WANG

[1,2,3] College of Computer Science, Beijing University of Technology, Beijing 100124, China

[4] Key Laboratory of Petroleum Resources Research, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China

Abstract - Fault recognition is an important section in seismic interpretation region and there have been many methods for this technology, but no one can recognize the fault exactly enough. For this problem, we proposed a new fault recognition method based on linear regression which can locate the position of a fault precisely and then extract it from the seismic section. First, the seismic horizons were labeled out by the eight adjacent connected component labeling method as connected regions; second, the horizontal endpoints of each connected component were found out based on the column coordinate of the pixels within the component; finally, the linear regression method based on least sum of square error was used to generate the direct line which was used to fit the horizontal endpoints. As a result, the direct line was regarded as the desired fault. To validate the availability and advancement of the proposed method, different fault recognition methods were compared through experiments on the synthetic seismic model data and the real seismic data. The comparison of the fault recognition results indicated that the proposed method is more accurate and effective than the traditional and latest presented methods.

Keywords - Fault Recognition, Labeling, Seismic Section, Linear Regression, Least Squares

1. Introduction

The recognition of the fault in seismic section plays an important role in seismic interpretation region. An accurate fault recognition result is very meaningful to oil and gas exploration work. In seismic section, horizon is the connected region which is continuous for a long distance in horizontal direction and fault is the part where the horizon is broken. Up to now, there is no good way to extract the fault from the seismic section directly, because they has no changeless property or shape, but the horizons have stable features such as coherence value within a specific window, horizontal gradient, variance and so on. Many researchers had used these properties for fault recognition, but the result was not accurate enough.

Conventional method to extract the fault from seismic data was to pick the points which were discontinuous with its neighbor sample in seismic section image by interpreter first, then these points were connected together manually to form the fault curve. This method can recognize the fault precisely, but it is time consumable for geological worker to pick each point manually, in addition, the manual method is too subjective because which point should be selected as fault may be not the same from different interpreter's opinion.

In order to overcome the inefficiency and strong subjectivity of the manual method, many automatic approaches have been proposed from its emergence at early twentieth century to now. The method that has been applied widely is the coherence cubes which have experienced three generations. The first generation was presented by Bahorich and Farmer (1996), both of them worked for the Amoco Corporation then. The method extracted the faults by the correlation value between three seismic traces based on the classic normalized cross correlation and it was referred to C1. The advantage of C1 is computationally simple and timesaving, but it easily disturbed by the coherence noise due to the number of the seismic traces it used is only three. In order to solve the problem the C1 algorithm has, Marfurt et al(1998) developed the second generation coherence cube, C2 for short, they expanded the correlative computation of three seismic traces to multi-traces contained in a cuboid or ellipsoid, this method have stronger capacity of resisting disturbance and higher recognition rate for big faults, besides, the signal-to-noise ratio of the results is higher than C1, but, the method has an obvious mean effect which made lots of small faults disappeared and the computation is much greater than C1 for the traces that used were much more. To speed up the C2, Gersztenkom and Marfurt(1999) proposed the third generation coherence cube, C3 for short, firstly, compute the eigenvalues of the covariance matrix of the seismic traces contained in the selected correlation window, then, on the basis of the variety of the seismic data was decided by the principal component which can be represent by the bigger eigenvalue, the method used the ratio of the biggest eigenvalue and the sum of all the eigenvalues as coherence value. The C3 was faster than C2 and its ability to resist noise is better, meanwhile, the faults could be recognized well. As a result, C3 has been used extensively in seismic interpretation industry, such as the geological software Petrel and so on. Nevertheless, C3 has several defects also, such that the size of the coherence window need to be adjusted to adapt to certain seismic data and it couldn't position the faults exactly, the dip angle information wasn't took into account and so on, besides, the time consumption still need to be reduced further. In addition to the coherence cubes, other faults recognition methods have been proposed also. Trygve Randen et al(2001) presented a new method of automatic fault surfaces extraction from three dimension fault enhancing attributes based on ant colony optimization, but they didn't explain the procedure in detail; Pedersen et al(2002,2005) presented a high-level fault interpretation workflow using automatically extracted surfaces by the method Trygve Randen et al had presented and provided a more specific interpretation for it and this method have been integrated to the geological software Petrel also, but the extracted fault's reliability was not high enough, because many surfaces that belong to fault-like surfaces had been reserved, in addition, the method was time-consumable for computing with three-dimension attribute data.

Gibson et al(2003) used the coherence cube as the discontinuous property to determine the mathematical model of the fault, then, used the maximum reliability prior strategy to extract the fault; Tingdahl et al(2005)extracted the mixed attribute feature of the fault first, then computed the fault probability cube through artificial neural network for the faults recognition; Admasu(2006) used faults highlight technology and active contour tracking method to extract faults; Kadlec et al(2008) proposed the interactional faults recognition method based on level set. Panagiotakis et al(2011) proposed an curve structure automatic enhance method for fault extracting based on the rotation and scale invariant filter. Hashemi(2012)proposed a faults recognition method based on the seismic section's illumination in full bandwidth frequency, the illumination of various frequency was obtained by the fuzzy cluster of the seismic sequences which used the time-frequency(TF) representations based on minimum mean cross-entropy(MMCE) solution, Bessel kernel and generalized marginal page distribution as the input. Panagiotakis et al(2015)proposed a faults recognition method based on the automatic enhancement and identification of the linear patterns of geological fault structures. Browaeys(2010) proposed a fault recognition method based on the complex-valued correlation of the instantaneous phase between neighbouring seismic traces, in this way, the fault can be quickly extracted, but there are more horizons of information is left behind. Wang et al(2016) proposed a method based on directional complex-valued coherence attributes, however, it is time-consuming and some horizons are left over similarly.

For its great significance and the methods introduced above can't recognize the faults well enough, we presented a new seismic fault recognition method based on the linear regression based on least sum of square error, FRLR for short. In statistics, linear regression (2014, 2016) is an approach for modeling the relationship between a scalar dependent variable and one or more explanatory variables (or independent variables). Linear regression was the first type of regression analysis to be studied rigorously, and to be used extensively in practical applications. In this paper, we use the row coordinate of the discontinuous point as the independent variable and the column coordinate as the dependent variable, then calculate the linear relationship between the row and column coordinates of the discontinuous point of the horizon by the linear regression method based on the least square of the error, the straight line from the discontinuous point that satisfies the linear relationship obtained from the regression is the desired fault. In order to validate the method's availability and advancement, we compared it with the traditional methods through experiments.

2. Fault Recognition Based on linear Regression

2.1 Positive Horizon Extraction and Faults Position's Location

2.1.1 Positive Horizon Extraction Based On Connected Region Label

The procedure can be introduced as the following:

a) Read the seismic section data and transform it to binary image by which the value of the pixel that belong to the horizon whose amplitude are positive is one and the value of the pixel that belong to the horizon or faults whose amplitude are negative is zero, then, create a label matrix whose size is the same to the binary image. After that, scan the binary image from upper left pixel (first line, first row) to the bottom right pixel (last line, last row) by a line-wise.

b) If the pixel value is zero, skip it and scan the next pixel in the same row continually. If the pixel value is one and its label in the label matrix is smaller than two, then track the outline of the connected region the pixel belong to and distribute a label number two or bigger (the number is increasing for different connected region) along the trace until return to the first pixel's position, the label number was saved in the corresponding position in the label matrix.

c) In the tracking process of the new connected region's outline, write down the front contour point first, then find the next contour point in the current contour point's eight-neighbor begin with but not include the front contour point in clockwise, as a result, the first point whose value is one will be judged as the next contour point. If there is no one point whose value is one, come back to the last contour point, by this way the whole outline of one connected region could be labeled with the same number. The process can be showed in Fig. 1 as below:

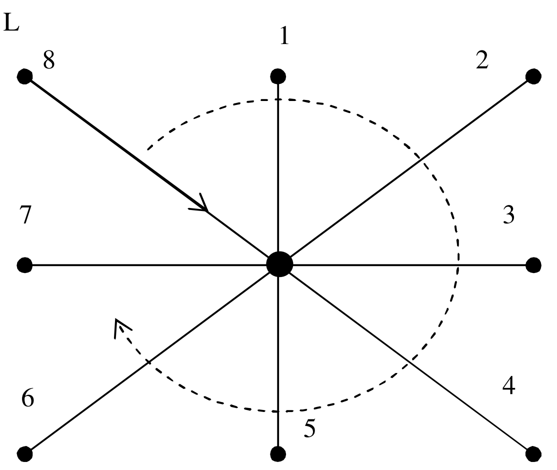

Fig. 1 Schematic diagram of the new connected region's outline tracking

Where, C is supposed as the current contour point and L is the last one, the next contour point will be searched in the points labeled 1 to 7.

d) During the scanning process, if the pixel value is one, then, check whether the corresponding label in the label matrix is smaller than two or not. If not, it indicates that the point has been labeled already; otherwise, it indicates the point has not been labeled yet, in this case, we will check whether there is a point in its eight-neighbor have been labeled or not, if there is one neighbor point have been labeled, then, the current point should be labeled with the same label to its neighbor, if there is no point in the eight-neighbor points have been labeled, the current point will be judged as a start point of a new connected region and the contour process begin with it, meanwhile, it will be labeled with a new label .

By the way described above, the positive horizons whose amplitude are bigger than zero could be extracted as the labeled connected regions and the other horizons whose amplitude are negative or zero was reserved with the faults.

2.1.2 Faults Position's Location

Because faults are the place where the horizon was not continuous, its position could be located precisely by horizontal endpoints of the positive horizons. The endpoint contains leftmost endpoint and right-most endpoint and they could be acquired according to the fellow equations:

$$(r_{left}, h_{left}) = \min_{L(r,h)=x} r \tag{1}$$

$$(r_{right}, h_{right}) = \max_{L(r,h)=x} r \tag{2}$$

Where, r_{left} and h_{left} represent the row and column coordinate of the leftmost endpoint respectively, L represents the label matrix, the objective function of the optimization is consist of the variable which represent the row-coordinate in L, $L(r,h)=x$ is the constraint condition which means that the variable need be limited to the points whose label is in L, the variable h is the column coordinate corresponding to r; The meaning of the variables and expressions in equation (2) can be inferred from the instructions for equation (1), the main difference between the two equations is that the objective function of the optimization (1) used is the minimization function while (2) used is the maximizing function.

Once each connected region's endpoints had been found out, the faults' position could be determined. But, we can't use the endpoints as the faults' position directly, because the endpoints belong to positive horizons not the faults. The method we adopted is use the point whose row coordinate is the same but its column coordinate is one unit smaller than the leftmost endpoint as the fault's position, likely, use the point whose column coordinate is one unit bigger than the right-most endpoint as the fault's position too. This can be explained by Fig. 2 as follows:

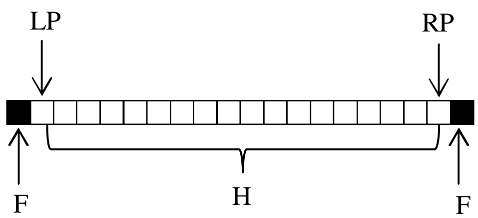

Fig. 2 Schematic diagram of the location of the fault's position based on the positive horizon's endpoints

Where, the capital H represents the positive horizon in the label matrix, LP and RP represent the left most endpoint and the right most endpoint respectively, both the capital F are regarded as the fault's position.

2.2 Fault recognition based on linear regression

The position of a fault has been located in previous, but they are discrete and the structure of the fault can't be well reflected by them. We use the linear regression method to determine the straight line with the least square of the error to the discontinuity points, linear least squares for short. The resulting straight line is regarded as a fault so that the overall structure of the fault can be reflected. A detailed description of the linear least squares method is given in the following.

Fitting requires a parametric model that relates the response data to the predictor data with one or more coefficients. The result of the fitting process is an estimate of the model coefficients. In the linear least squares method, the model coefficients consisted of slope and intercept of the desired line. Suppose the vectors $X=(x_1,x_2,\cdots,x_n)$ and $Y=(y_1,y_2,\cdots,y_n)$ are the independent variable and response value of the observed data points(n is the number of data points), the purpose of the linear least squares method is to estimate the slope and intercept of the equation for the fitted straight line whose equation is described in Equation (3).

$$\hat{Y} = p_1 \bullet X + p_2 \tag{3}$$

Where, $\hat{Y}=(\hat{y}_1,\hat{y}_2,\cdots,\hat{y}_n)$ is the vector of the predicted response values, p_1 and p_2 represent the slope and intercept of the fitted straight line, respectively.

To obtain the coefficient estimates, the least squares method minimizes the summed square of residuals. The residual for the ith data point is defined as the difference between the observed response value y_i and the fitted response value $\hat{y}_i$, and is identified as the error associated with the data. The residual r_i for the ith data point and the summed square of residuals S is given by Equation (4) and Equation (5) as follows.

$$r_i = (\hat{y}_i - y_i)^2 \tag{4}$$

$$S = \left|\hat{Y} - Y\right| = \sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{5}$$

Based on the above analysis, the process of linear least squares method can be represents by Equation (6) to Equation (8) as bellow.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \times \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \tag{6}$$

$$P = \left(X^T X\right)^{-1} X^T Y \tag{7}$$

$$\hat{Y} = X\left(X^T X\right)^{-1} X^T \tag{8}$$

Where, $P=[p_1,p_2]^T$, the italic letter " T " denotes a matrix transpose operation. The result $\hat{Y}$ in equation (8) is the predicted value obtained by linear regression. An example of a linear regression is shown in Fig. 3.

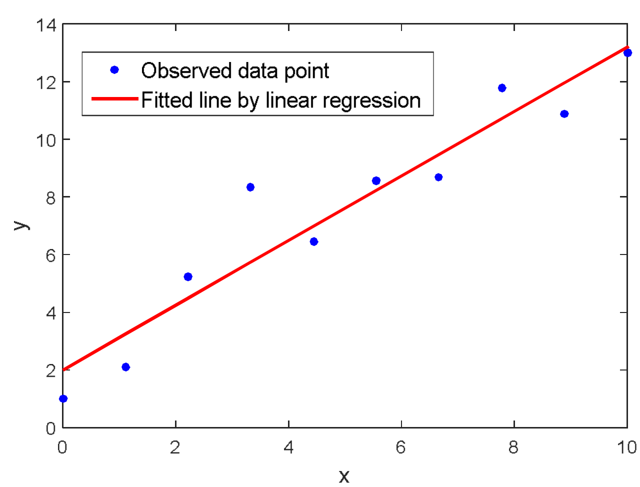

Fig. 3 Example of the linear regression

In this paper, the row and column coordinates of the fault position are used as independent and response values of the observed data point respectively, the resulting straight line fitted by the linear least squares method is regarded as the desired fault.

3. Experiments and Comparing

In order to validate the effectiveness of the FRLR proposed in this paper, we did experiment based on the seismic model data and real seismic data. The size of the seismic model data was 301×151 pixels, the size of the real seismic data was 220×140 pixels. The experiment was carried in Matlab2015a on the PC whose operating system was windows 7(64 bit) and the size of the internal memory was 4G. The traditional methods used to compare with the FRLR include the C3 method, the method proposed by Browaeys(2010), SEG2010 for short, and the method proposed by Wang et al(2016), JAG2016 for short. The comparison of the fault recognition results and the time consumption of different methods are given bellow.

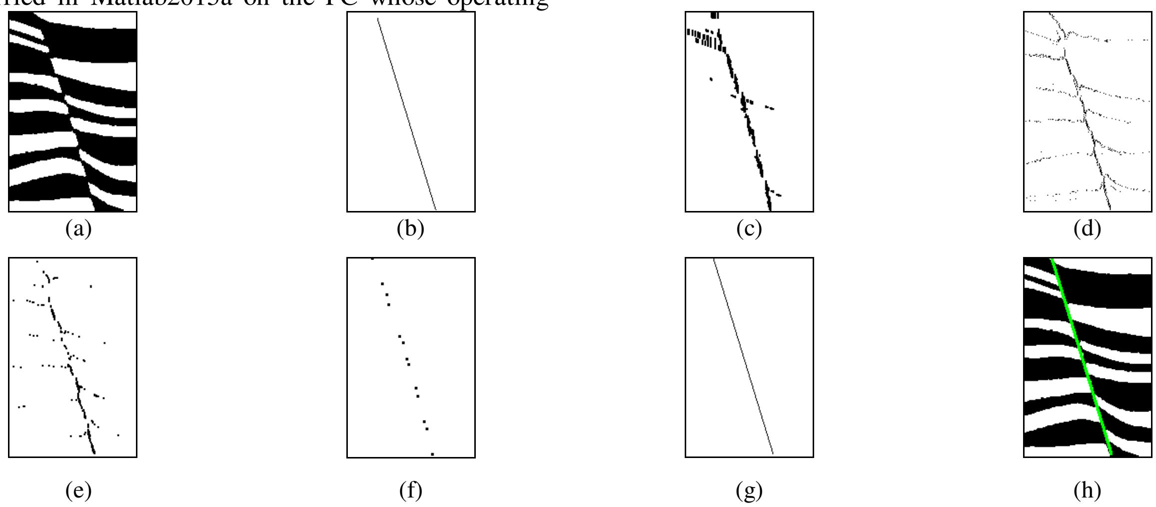

Fig. 4 Fault recognition comparison of the traditional method and the proposed method on the seismic model data

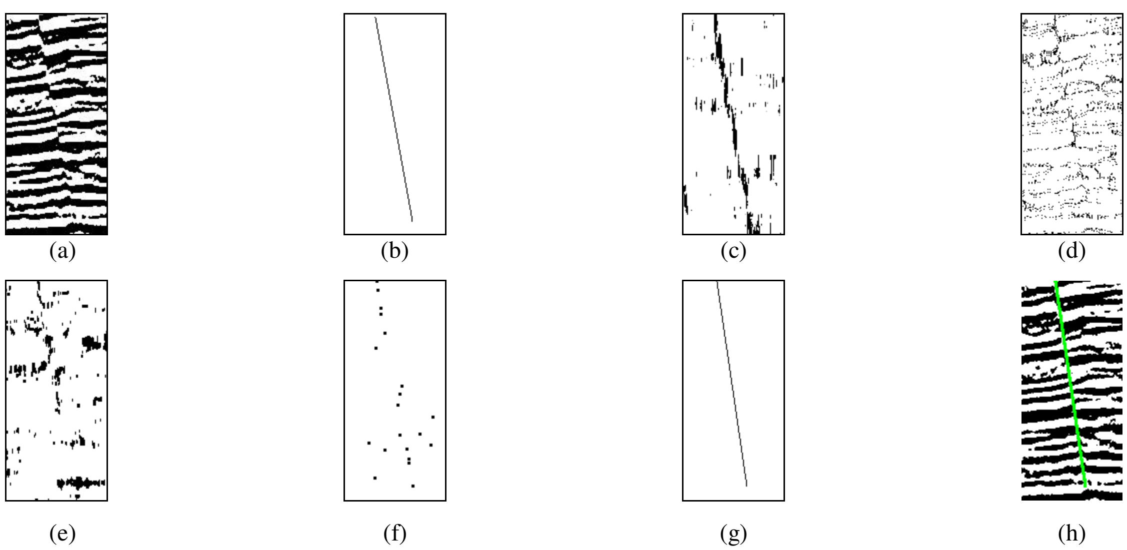

Fig. 5 Fault recognition comparison of the traditional method and the proposed method on the 82th real seismic section

Fig.4 is the comparison of the fault recognition results of the traditional method and the FRLR on the seismic model data, Fig.5 is the comparison of the fault recognition results on the 82th real seismic section. In Fig. 4 to Fig. 5, (a) is the original image of the seismic section (model or real seismic section), (b) is the fault recognition result of manual method, (c) is the fault recognition result of C3, (d) is the fault recognition result of SEG2010, (e) is the fault recognition result of JAG2016, (f) is horizon's endpoints results determined by the connected region label method, (g) is the fault recognition result by the FRLR method, (h) is the final result in which the extracted fault and the original seismic section were combined together.

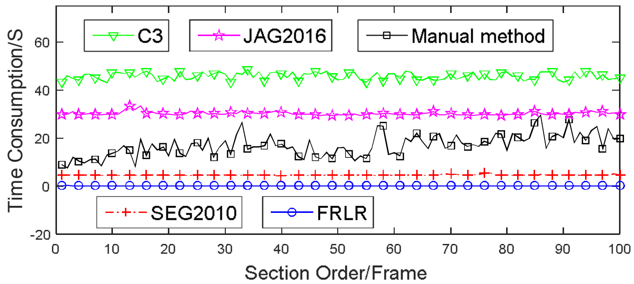

Fig. 6 Comparison of time consumption between the traditional methods and FRLR in each seismic section

From the comparison of the fault recognition results of the traditional methods and the FRLR method proposed in this paper, we could see that the result of the FRLR method is more accurate than the traditional methods and the effectiveness of the FRLR algorithm could be certified. In addition, by the time consumption comparison between the traditional methods and FRLR in each seismic section as showed in Fig. 6, we could see that, the FRLR method is the fastest one and its real-time performance could be certified also. The average time consumption comparison can be showed in Table 1 as below:

Table 1 Average time consumption comparison

Method	Average time consumption(s)
Manual method	16.9589
C3	45.8452
SEG2010	4.6014
JAG2016	30.2568
FRLR	0.1599

From the comparison described above, we could know that the FRLR method proposed in this paper is superior to the traditional ones not only in the aspect of effectiveness but in real-time performance also.

4. Conclusions

To recognize the fault exactly enough, we proposed a new fault recognition method based on linear regression method in this paper. At the beginning, the method of determining fault location is introduced, then, the principle of the linear regression method based on least squares fitting and the proposed fault recognition method based on the linear regression were given out; finally, to validate the availability and advancement of the proposed method, different fault recognition methods were compared through experiments on the synthetic seismic model data and the real seismic data. The experimental results indicated that the proposed method is more accurate and effective than the traditional presented methods. To apply the proposed method to the problem of multiple faults recognition is our future work.

Acknowledgments

All the work we have done was supported by National Natural Science Foundation of China(61501008)and Beijing Natural Science Foundation (4162007).

References

[1] Bahorich M S, Lopez J, Haskell N L, Nissen S E, Poole A. Stratigraphic and structural interpretation with 3-D coherence. In: Proceedings of the 65th Annual International SEG Meeting, Houston, Expanded Abstracts, 1995.97-100

[2] Marfurt K J, Kirlin R L, Farmer S L, Bahorich M S. 3-D seismic attributes using a semblance-based coherency algorithm. GeoPhysics, 1998, 63(4):1150-1165

[3] Gersztenkorn A, Marfurt K J. Eigenstructure-based coherence computations as an aid to 3-D structural and stratigraphic mapping. GEOPHYSICS, 64(5):1468–1479

[4] Trygve Randen, Stein Inge Pedersen, Lars Sønneland. Automatic Extraction of Fault Surfaces from Three-Dimensional Seismic Data[C]. In: Proceedings of the 62nd Ainnual International SEG Meeting, San Antonio, Texas, SEG Technical Program Expanded Abstracts 2001: 551-554.

[5] Pedersen S I, Randen T, Sønneland L and Steen ø.. Automatic fault extraction using artificial ants. Expanded Abstracts of the 71th Annual International SEG Meeting, 2002, 512-515

[6] Pedersen S.I， Skov T， Randen T, Sønneland L. Automatic fault extraction using artificial ants[J]. Springer Berlin Heidelberg, 2005, 7:107-116

[7] Gibson D, Spann M, Turner J. Automatic fault detection for 3D seismic data. In: Proceedings of the VIIth Digital Image Computing: Techniques and Applications, Los Alamitos: IEEE Computer Society Press, 2003. 821-830

[8] Tingdahl K M, Rooij M D. Semi-antomatic detection of faults in 3D seismic data. GeoPhysical Prospecting, 2005, 53, 533-542

[9] Admasu F, Back S, Toennies K. Autotracking of faults on 3D seismic data. GeoPhysics, 2006, 71(6):A49-A53

[10] Kadlec B J, Dorn G A, Tufo H M, Yuen D A. Interactive 3-D computation of fault surfaces using level sets.Visual Geosciences, 2008, 13(l): 133-138

[11] Panagiotakis C, Kokinou E, Sarris A. Curvilinear structure enhancement and detection in geophysical images. IEEE transactions on geoscience and remote sensing, 2011, 49(6): 2040-2048

[12] Hashemi H. Fuzzy clustering of seismic sequences: segmentation of time-frequency representations. IEEE Signal Processing Magazine, 2012, 29(3): 82-87

[14] Panagiotakis C, Kokinou E. Linear pattern detection of geological faults via a topology and shape optimization method. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015, 8(1): 3-11.

[15] Browaeys T J. Local complex-valued correlation of seismic phases. In: Proceedings of the 80th Annual International SEG Meeting (SEG2010). Denver, Colorado, USA, Expanded Abstracts, 2010: 1423-1427.

[16] Wang, S.X.; Yuan, S.Y.; Yan, B.P.; He, Y.X.; Sun, W.J. Directional complex-valued coherence attributes for discontinuous edge detection. Journal of Applied Geophysics. 2016, 129, 1-7.

[17] Feng Lu; Yusuke Sugano; Takahiro Okabe; Yoichi Sato. Adaptive Linear Regression for Appearance-Based Gaze Estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(10): 2033–2046.

[18] Ting Hu; Qiang Wu; Ding-Xuan Zhou. Convergence of Gradient Descent for Minimum Error Entropy Principle in Linear Regression. IEEE Transactions on Signal Processing, 2016, 64(24): 6571–6579.

Authors -

Lei CHEN received his Master. degree in Pattern recognition and intelligent system from Taiyuan University of Science and Technology, Taiyuan China. He is currently a PhD Scholar at the College of Computer Science and Technology in Beijing University of Technology. His research interest is pattern recognition and computer vision.
BOD: 1981-04-10.

Chuang-Bai XIAO Professor at the School of Computer Science and Technology, Beijing University of Technology, His main research interest is digital signal processing.
BOD: 1962.

Jing YU lecturer at the School of Computer Science and Technology, Beijing University of Technology, Her main research interest are Image processing , pattern recognition.
BOD: 1981.

WANG Zhen-Li Director and associate professor at the Key Laboratory of Petroleum Resources Research, Institute of Geology and Geophysics, Chinese Academy of Sciences. His main research interest is Complex structure imaging.
BOD: 1963