LARS Technical Report 080982

Detection of Bright Spots Using Pattern Recognition Techniques

Kou~Yuan Huang, Clare D. McGillem, and Paul E. Anuta

School of Electrical Engineering
and
Laboratory for Applications of Remote Sensing
Purdue University West Lafayette, Indiana 47907 USA
1982

DETECTION OF BRIGHT SPOTS USING PATTERN RECOGNITION TECHNIQUES

Kou-Yuan Huang, Clare D. McGillem, and Paul E. Anuta

Laboratory for Applications of Remote Sensing
and School of Electrical Engineering

Purdue University

West Lafayette, IN 47907, U.S.A.

ABSTRACT

Successful application of automatic information extraction techniques to seismic exploration data would greatly relieve the burden of manual inspection of large quantities of seismic section plots. We describe here work directed toward generation of features and classification methods for pattern recognition of seismic signals. In this work, a simulated seismogram of primary reflections is generated for a simple bright spot model. The first task discussed is the estimation of the reflection coefficients from layer interfaces in the model. Gaussian noise with a negative exponential autocorrelation function was added to the synthetic seismogram to simulate reverberations, weak reflections, and instrumentation noise.

From the simulated seismogram, the density profile is estimated. The velocity distribution is assumed to be known from CDP stacking analysis.

Presented at Society of Exploration Geophysicists 51st annual international meeting and exposition, Los Angeles, October 1981.

The reflection coefficients are estimated from the polarity and amplitude of the reflected signal as derived from the analytic signal representation. The reflecting interfaces are assumed to be located at the times of peaks of the envelope of the analytic signal. From these calculations, a sequence of estimated reflection coefficients is found. Densities of the layers between each interface are computed from the expression for normal incidence reflection coefficient. The densities are computed iteratively, assuming the first layer density is known. The velocity values from the bright spot model are modified with noise to simulate errors in the CDP analysis process. Statistical pattern analysis is then performed on the two variables to discriminate the different layers in the model. Maximum likelihood classification of the velocity and density values estimated for each layer are performed and results are presented for several signal-to-noise ratios.

INTRODUCTION and GENERATION OF SYNTHETIC SEISMOGRAM

Seismic petroleum exploration has become more and more important in recent years and will continue so in the future as the energy crisis becomes more severe. Because of this, there is an ever-increasing load on the data analyst and there is, therefore, a need to consider new data processing procedures. One of the data analysis techniques that has not yet found application in seismic exploration is that of pattern recognition. In this paper, the application of this technique to a simple analysis problem is described.

The problem to be considered is that of a bright spot model as shown in Figure 1 which consists of six distinct layers. The density and velocity are assumed constant throughout each layer although in practice they would be expected to vary with depth. In the gas and oil zones, the densities and velocities are significantly lower than those in the other zones. The gas sand zone has density D=1.8 gm/cm**3 and velocity V=1.8km/sec. And the oil sand zone has density D=2.2 gm/cm**3 and velocity V=2.2 km/sec.

The amplitude of a seismic wave reflected from an interface between two materials is governed by the reflection coefficient R which is expressed for normal incidence by the relation

$$R=(D(2)*V(2) - D(1)*V(1))/(D(2)*V(2) + D(1)*V(1))$$

where D(1) and D(2) are the respective densities on the near (incident) and far sides of the boundary and V(1) and V(2) are the respective velocities for the two sides. The product of D and V is known as the acoustic impedence. A high-amplitude portion of a seismic trace corresponding to a high reflection coefficient is referred to as a bright spot[4].

Convolution of the average excitation wavelet W(t) with a system function R(t) consisting of delta functions at the boundary layers having amplitudes equal to the primary reflection coefficients generates the seismogram S(t) of bright spots as shown in Figure 2. The seismogram plus zero mean Gaussian random noise with variance 0.015^2 is shown in Figure 3. The mathematical expression for the seismic signal plus noise is

$$S(t) = R(t) * W(t) + N(t)$$

The average wavelet W(t) was chosen to be a duration-limited 25 Hz Ricker wavelet. The minimum reflection coefficient is 0.11 corresponding to a signal-to-noise ratio of 17.3 dB. A sampling rate of 0.004 seconds/data interval is employed. The model contains 512 seismic traces. The total receiving time is 2.044 seconds, corresponding to 512 points in each trace (line). Figure 3 consists of 512x512 pixels. The value of the pixel is the seismogram value scaled from 0 to 255. The data of Figure 3 will be used in the pattern recognition procedure.

SEISMIC PATTERN RECOGNITION SEQUENCE

The seismic pattern recognition approach is shown in Figure 4. The simulated seismogram of Figure 3 is assumed to correspond to the stacked seismogram. The reflection seismogram will be transformed into density and velocity data. From the CDP stacking technique, the optimal stacking velocity can be derived[1]

PREPROCESSING: CALCULATION OF REFLECTION COEFFICIENTS FROM THE ANALYTICAL SIGNAL

Reflection coefficients are calculated from the data of Figure 3 by the procedure shown in Figure 5.

The envelope of the signal is calculated as the square root of the sum of the square of the signal S(t) and its Hilbert transform S(t)[2,5]; thus

$$\text{Envelope}\left\{S(t)\right\} = \sqrt{S(t)^{**}2 + S(t)^{**}2}$$

The envelope describes the outer shape of the wavelet, even in the noise case. The envelope of Figure 3 is shown in Figure 6.

Envelope values below an appropriate threshold level are set to zero. The threshold is selected empirically to reduce the effect of noise. The next step is to find the maxima of the envelope function. Because the reflection coefficient between the free surface and the first layer is unity, the amplitudes can be compared to this value to determine the corresponding reflection coefficients.

The "peak signal" is defined as the value of the normalized maximum with the polarity of the original signal at that point. The "peak signal" is made up of the reflection coefficients R(1), R(2),...,R(n-1) at the boundaries.

FEATURE SELECTION

Two features sufficient for classifying the various strata are the density and velocity at each point. These can be estimated from the seis-

mogram. From the CDP stacking velocity analysis, one feature--velocity distribution--is found. The interval velocity of the same layer is assumed constant along one trace (line) and to have a Gaussian random distribution for different traces (lines), as shown in Figure 7. The other feature is density. It is assumed that the first layer's density D(1) is known and that the reflection coefficients R(1), R(2),...,R(n-1) can be calculated from "peak signal." Then for normal incident, the reflection coefficient is

$$R(1) = (D(2) * V(2) - D(1) * V(1)) / (D(2) * V(2) + D(1) * V(1))$$

Solving for the density D(2) gives

$$D(2) = (D(1) * V(1) / V(2)) * ((1+R(1) / (1-R(1))$$

Repeating this calculation for subsequent layers gives D(2), D(3), D(4),...,D(n). The result of this calculation is shown in Figure 8. Since density is calculated iteratively, there is error propagation. The deeper the layer, the larger the error, as shown in Figure 9. From Figure 8 it is seen that at the junctions of the layers (shale and gas, shale and oil), the reflections from the boundaries are mixed. The reflections of the two layers cannot be distinguished, since the two boundaries are too close. This is called a disturbance area. It affects calculation of the reflection coefficients from the seismogram and affects the density calculation. The number of layers can be reduced or increased by computer calculation, as shown in Figure 8.

PROCESSING OF SIMULATED EXPLORATION SEISMIC PATTERN RECOGNITION

A seismic exploration supervised classification analysis flow chart is shown in Figure 10.

In order to use an existing data analysis and classification system[3], the sample values are adjusted to cover the dynamic range of (0, 255) from their actual range of (1.4, 3.3). Any data value less than 1.4 is assigned to 0 and any data value over 3.3 is assigned to 255.

For each trace (line), the density or velocity is assumed to be the same value in the same layer, but they are random for different traces (lines) in the same layer. A set of training areas in the seismogram is selected which is representative of the various classes (layers) present. The training areas are from trace 210 and 290 and one horizontal line for one layer because the data are redundant. The testing area is from trace 150 to 350 and one horizontal line for one layer. The classifier computes the statistics of these classes from the training sets and then goes through the test and the entire data sets and assigns each point to that class for which it has the greatest probability. The result is shown in Figure 11. The statistical mean of the density and velocity at the selected training area can be linked with the kinds of geology. The computed means are D(1)=1.596 gm/cm**3, V(1)=1.596 km/sec, D(2)=1.963, V(2)=1.997, D(3)=2.420, V(3)=2.495, D(4)=1.709, V(4)=1.797, D(5)=2.063, V(5)=2.196, D(6)=2.585, V(6)=2.797. The training and test areas avoided the disturbance areas. The supervised classification based on the maximum likelihood decision rule is used. For a given sample X, if P(X/W1) >

P(X/W2), then X belongs to W1 class. If P(X/Wi), i=1,2,...n, is a multivariate Gaussian density function. The classifier is

$$1/2 \log \frac{|Ki|}{|Kj|} - 1/2[(X-M_i)^T Ki^{-1}(X-M_j) - X-M_j)^T K_j^{-1} (X-M_j)] = 0$$

The classifier is quadratic if covariance matrices are different.

The performance is 100% in the above simulation. Because the probability distribution of density and velocity does not overlap in Figure 9, the classification error is zero and the gas and oil layers are clearly delineated.

The confusion matrix P is an identity matrix from 100% performance,

For $Z = P^{-1} \hat{Z}$

so $Z = \hat{Z}$

where $Z = \begin{bmatrix} V \\ D \end{bmatrix}$ is the matrix of the actual number pixels of classes,

$\hat{Z} = \begin{bmatrix} \hat{V} \\ \hat{D} \end{bmatrix}$ is the matrix of the classified number pixels of the classes.

Although there are serious distortions in density distribution at the disturbance areas, the assumed velocity has no serious distortion in these areas. This was sufficient to permit 100% classification performance to be obtained.

For the same processes, the classification result for a signal-to-noise ratio of 15.7 dB (noise $\sigma^2=0.018^2$) is shown in Figure 12.

CONCLUSIONS AND SUGGESTIONS

Conclusions:

1. Two features, density and velocity, can be estimated from seismogram. Velocity is obtained from seismic CDP stacking velocity analysis. Density D(1) is known, then D(2), D(3),..., D(n) are calculated from iteration by D(2) = (D(1) * V(1) / V(2)) * ((1+R) / (1-R)).

2. Because the calculation of density is iterative, errors will propagate. The distribution of density in Figure 9 is larger than the velocity distribution at every layer. The deeper the layer, the greater the distribution of the density because of error propagation.

3. Use of the envelope can improve the calculation of the reflection coefficients because it describes the outer shape of the wavelet, even in the noise case.

4. Supervised classification using the maximum likelihood decision rule was highly accurate in this simulation at the chosen S/N ratios. Gas and oil layers are clearly classified, although there are serious distortions in density distribution at the junction of shale and gas layers and the junction of shale and oil layers. Lower signal-to-noise ratios were not tested in this study.

Suggestions:

1. Change the assumed constant stacking velocity at every layer to other velocity functions. For example, use linear function V = V0+A*Depth, where A is a constant. That will approach the real case.

2. Add larger noise to the velocity distribution, then density and velocity distribution will overlap in this simulation. The classification accuracy will approach what may be the encountered in practice.

3. Thinner gas and oil layers can be assumed.

4. Find other features for pattern recognition; for example, the well log information.

5. Deconvolution techniques may improve the result at the disturbance areas.

REFERENCES

[1] Taner, M. Turhan and Fulton Koehler. Velocity Spectral Digital Computer Derivation and Applications of Velocity Functions. Geophysics, Vol. 36, 1969, pp. 859-881.

[2] Huang, K.Y., C.D. McGillem, and P.E. Anuta. Analytic Signal Representation in the Synthetic Seismogram of Bright Spots. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, April 1, 1981, pp. 1062-1065.

[3] LARS Staff, R.P. Mroczynski, ed. LARSFRIS User's Manual, Laboratory for Applications of Remote Sensing (LARS), Purdue University, West Lafayette, IN 47906-1399. Oct. 1980. (Six volumes.)

[4] Dobrin, Milton B. Introduction to Geophysical Prospecting, third edition. McGraw-Hill, 1976, Chapter 10.

[5] Taner, M.Turhan, Fulton Koehler, and R.E. Sheriff. Complex Seismic Trace Analysis. Geophysics, Vol. 44, 1979, pp. 1041-1063.

FIGURES

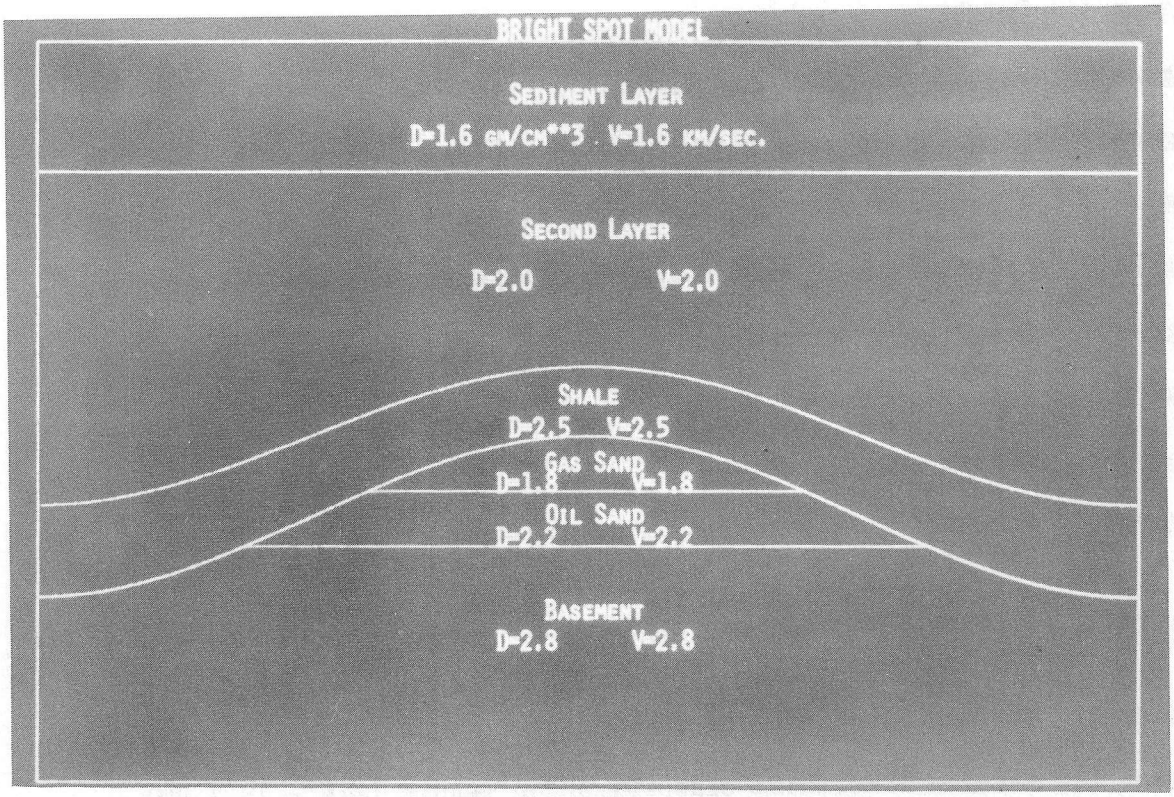

Figure 1. Bright spot model.

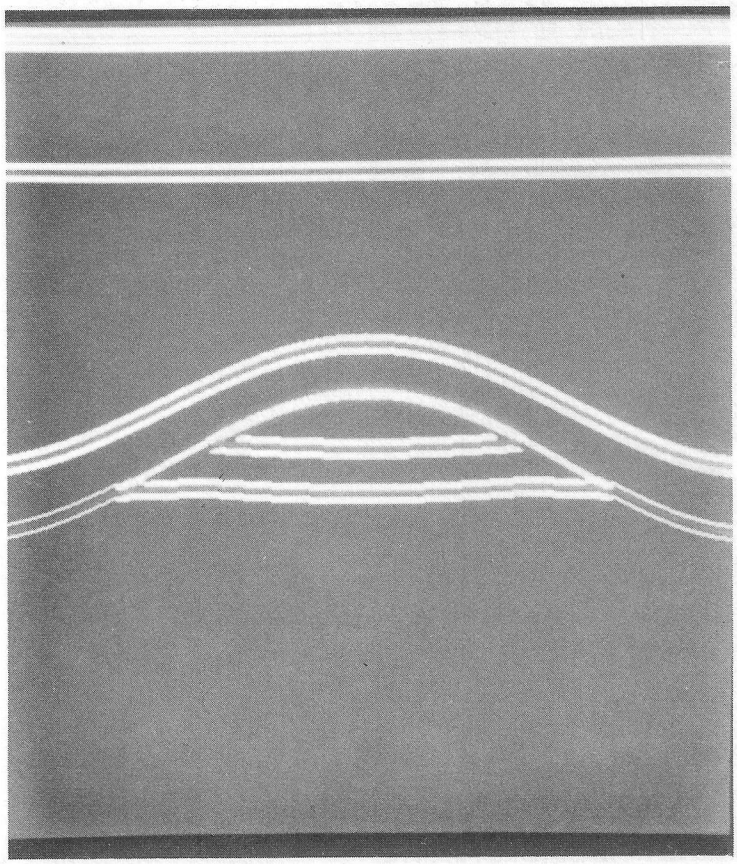

Figure 2. Seismogram of bright spot.

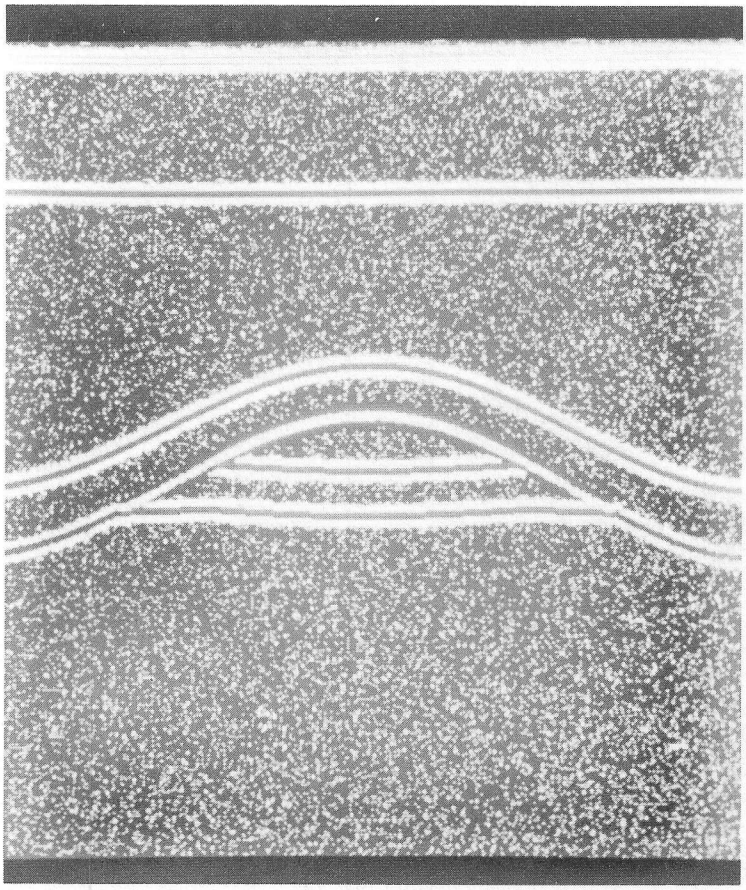

Figure 3. Signal plus zero-mean Gaussian random noise (S/N = 17.3 dB).

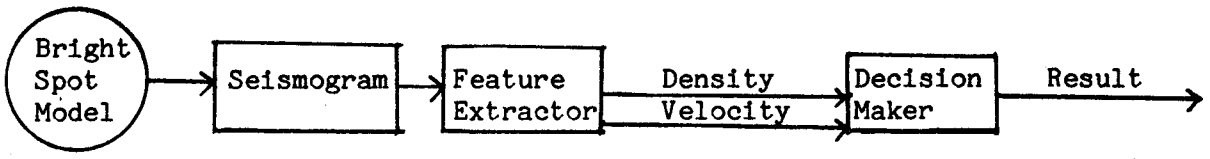

Figure 4. A seismic pattern recognition sequence.

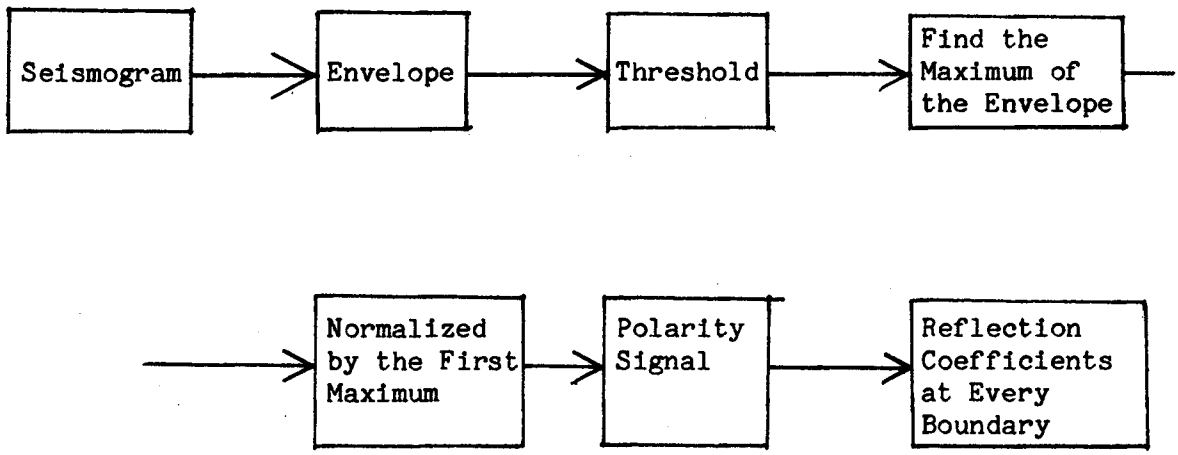

Figure 5. Calculation of reflection coefficients from analytical signal.

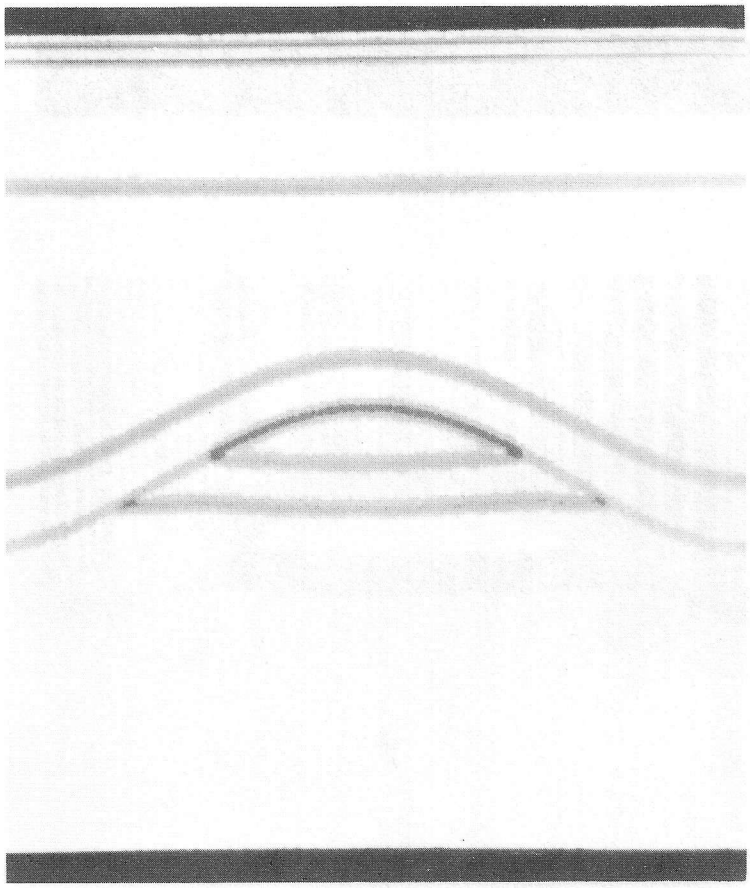

Figure 6. Envelope of Figure 3.

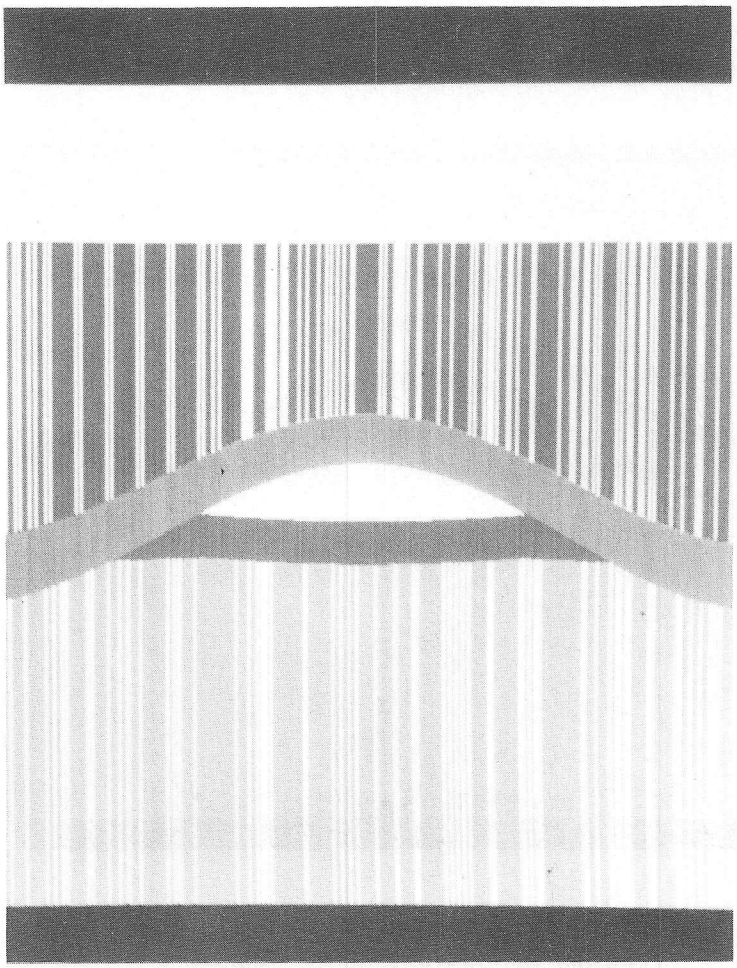

Figure 7. Velocity distribution.

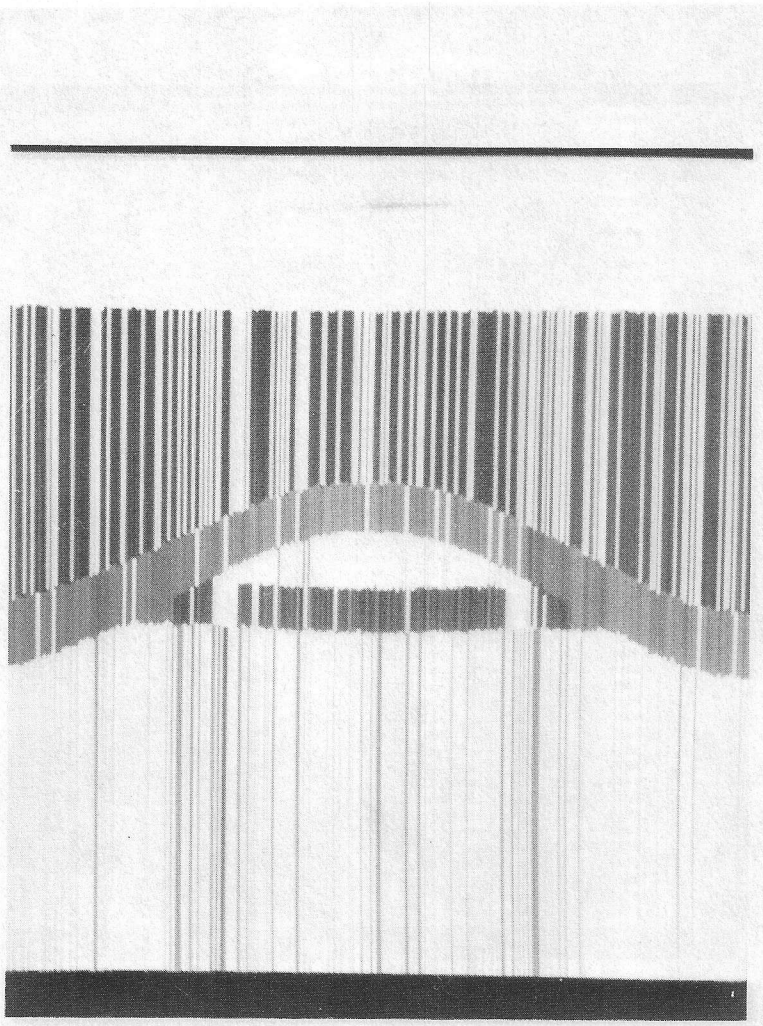

Figure 8. Density distribution.

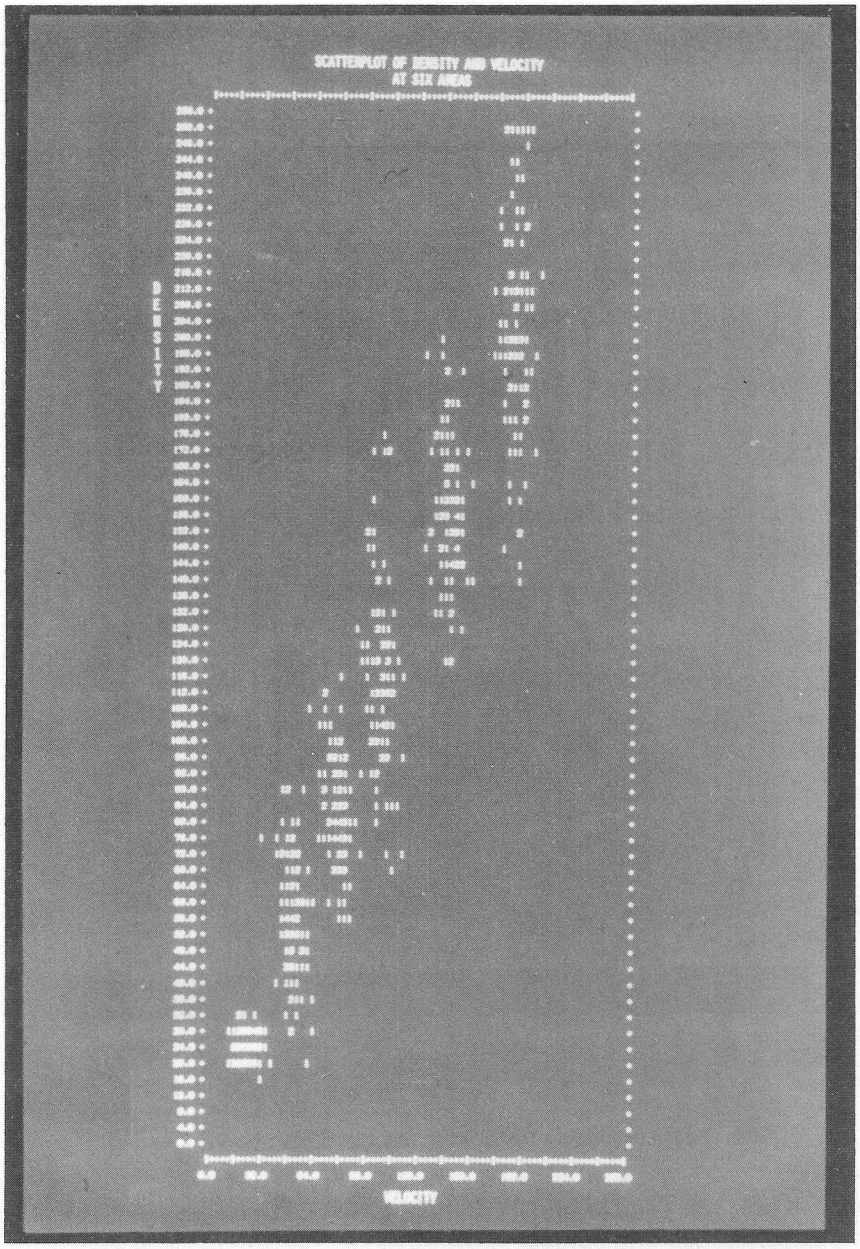

Figure 9. Velocity versus density from training areas of six layers (Trace 210 to 290).

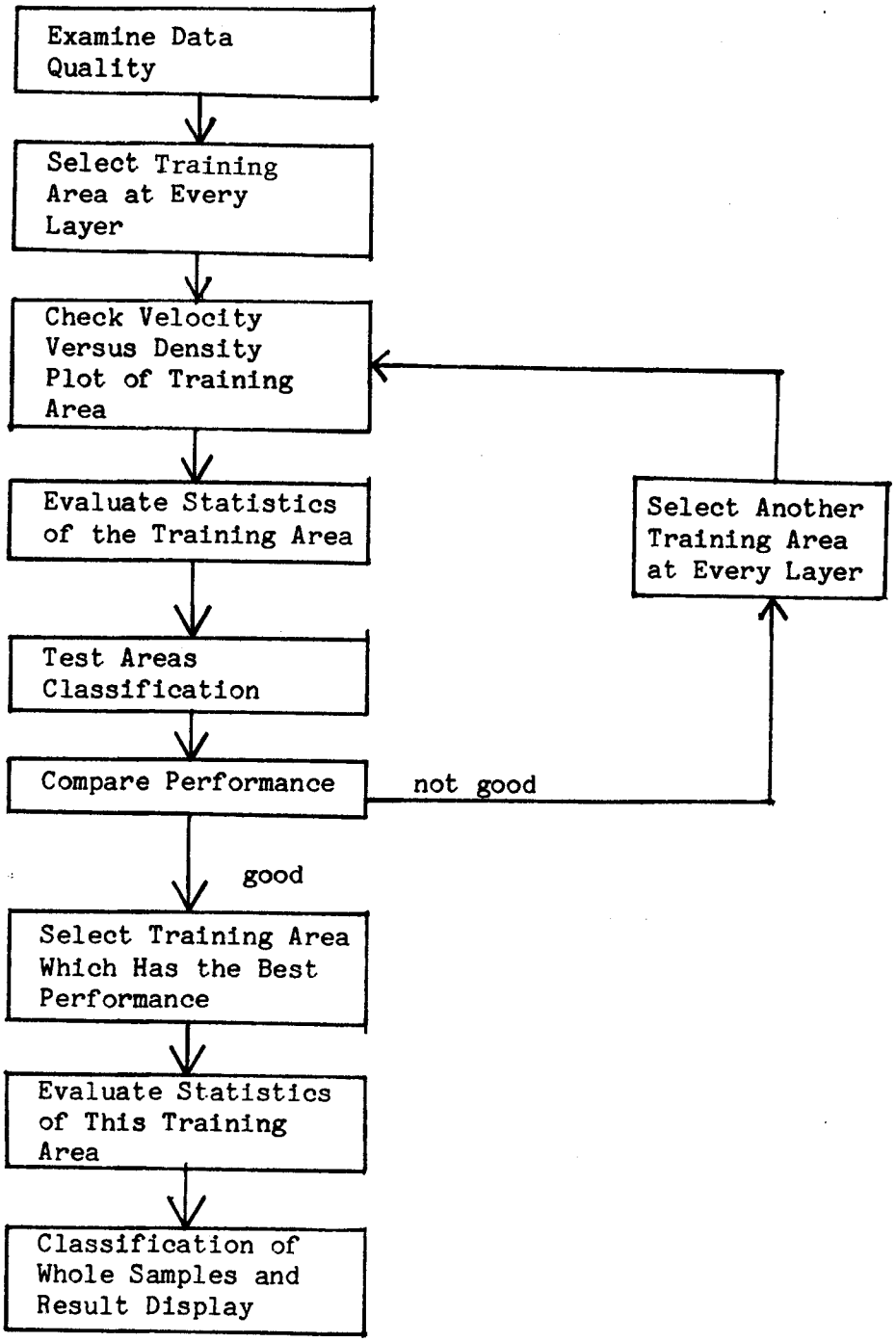

Figure 10. A supervised seismic exploration classification analysis flow chart.

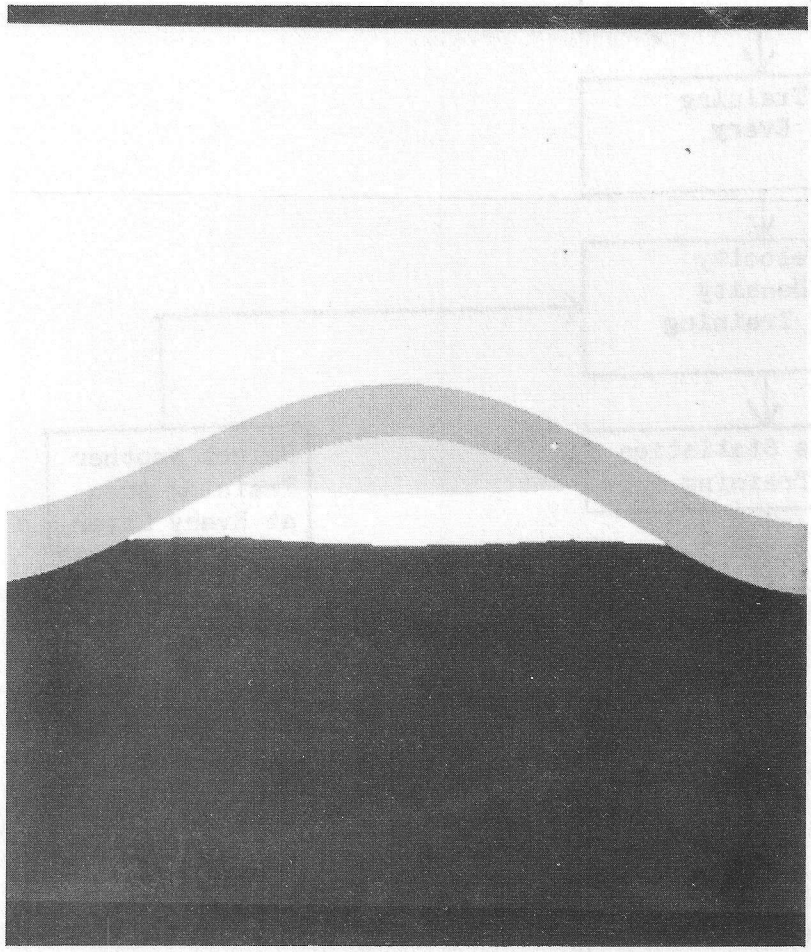

Figure 11. Maximum likelihood classification result (S/N = 17.3 dB).

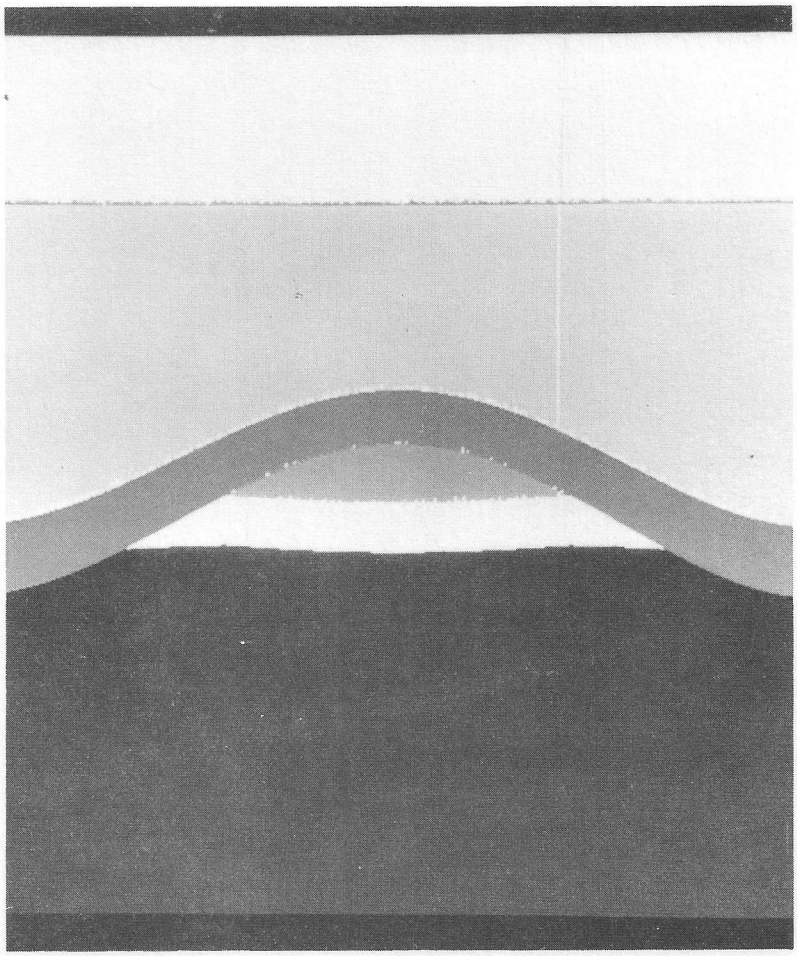

Figure 12. Maximum likelihood classification result (S/N = 15.7 dB).