

4 Rastreamento de horizontes e falhas

Os horizontes sísmicos e as falhas sísmicas são dois eventos sísmicos importantes, sendo relevante encontrar sua localização ao interpretar dados sísmicos.

Uma das técnicas mais importantes do processo de interpretação de dados sísmicos consiste no rastreamento dos horizontes. Um horizonte sísmico é definido na geologia como uma subsuperfície presente na região geológica onde os dados foram adquiridos. O rastreamento de um horizonte sísmico consiste em identificar em quais dos traços do volume o horizonte desejado está presente, identificando as amostras que pertencem ao horizonte em questão.

As características dos horizontes retidas nos volumes sísmicos permitem que um horizonte seja identificado nos traços por um padrão de amplitudes encontrado na sua vizinhança vertical, padrão esse que se repete ao longo dos traços do volume onde tal horizonte está definido.

Horizontes reais podem ser representados nos traços do volume por um pequeno conjunto de voxels verticalmente contíguos. No entanto, na nossa modelagem iremos considerar que um horizonte é representado por no máximo um voxel por traço do volume. Essa consideração facilita bastante a modelagem do problema.

A maioria dos métodos desenvolvidos para o mapeamento de horizontes possui algumas limitações indesejáveis. Por serem métodos baseados em uma estratégia de crescimento de região (*Region Growing*), esses métodos adicionam novos voxels da superfície a partir dos voxels já descobertos do horizonte nos traços vizinhos (para o início do processo de extração, uma semente é selecionada externamente). Tais métodos conseguem bom desempenho em volumes onde não existe a presença de falhas sísmicas ou onde, apesar da existência de tais falhas, o horizonte a ser mapeado forma uma superfície conexa no volume. No entanto é bastante comum encontrar horizontes para os quais a porção conexa da superfície

não está capturada nos dados, o que fará com que a procura por vizinhos aos voxels descobertos cesse nas regiões delimitadas pelas falhas, comprometendo o resultado final obtido.

Nos algoritmos de extração de horizontes a partir de dados sísmicos, a informação de contexto espacial não pode ser descartada. Voxels de um mesmo horizonte são fortemente caracterizados por sua similaridade considerando sua própria amplitude em contraste com as intensidades da sua vizinhança no traço sísmico. Desse modo, um método de agrupamento de dados em que os voxels de um mesmo horizonte sejam representados no conjunto de amostras não somente por sua amplitude, mas também pelas amplitudes da sua vizinhança vertical, tenderia a situar voxels de um mesmo horizonte em um mesmo grupo ou em grupos situados próximos entre si no espaço das amostras.

Além dos horizontes, outra estrutura geológica importante a ser identificada durante o processo de interpretação dos dados são as falhas sísmicas. Essas estruturas são quebras na continuidade original dos horizontes. São fraturas que causam um deslocamento relativo das rochas, fazendo com que elas percam sua continuidade original. Nos dados sísmicos, os voxels pertencentes às falhas são caracterizados por sua baixa coerência em relação a sua vizinhança. O mapeamento de uma falha sísmica consiste em encontrar o conjunto de voxels do volume que evidencia essa quebra de continuidade. Ao interpretar um conjunto de dados, o mapeamento das possíveis falhas sísmicas presentes no dado é extremamente importante.

Neste capítulo são apresentados algoritmos para a extração de horizontes sísmicos e para o mapeamento de falhas sísmicas, ambos baseados em uma estratégia de agrupamento de dados na qual os voxels do volume são considerados no contexto de sua vizinhança vertical. No início do processo, para cada voxel do volume é criado um vetor de características que retém essa vizinhança. É esse vetor que irá representar o voxel e sua vizinhança no conjunto de amostras de entrada utilizado no processo de treinamento do algoritmo GNG. O capítulo está organizado da seguinte maneira: na seção 4.1 é descrito em detalhes o processo de criação do conjunto de dados de entrada a serem utilizados na etapa de treinamento do algoritmo de GNG. Na seção 4.2 discutimos as funções de similaridade testadas durante o processo de criação do grafo. Uma interpretação

geométrica do processo de agrupamento das amostras em relação à sua classificação nos neurônios é apresentada na seção 4.3, ficando o processo de extração dos horizontes a partir do volume quantizado descrito na seção 4.4. Na seção 4.5 é discutido um método para extração de falhas sísmicas, que apesar de ainda estar em fase inicial de pesquisa apresenta resultados que sugerem que a técnica pode ser uma boa alternativa para o mapeamento automático desse tipo de dados.

4.1 Criação do conjunto de amostras de entrada

Para reter a informação contextual vertical de cada voxel, ao criar as amostras o i -ésimo voxel do traço t , v_{ti} (um escalar) é representado no conjunto de amostras de entrada por um vetor s_{ti} contendo o valor da sua amplitude na posição central da amostra, além das amplitudes dos seus k vizinhos acima e abaixo no traço sísmico, fazendo com que cada uma das amostras seja formada por um total de $2k+1$ voxels:

$$S_{ti} = \{v_{t(i-k)}, \dots, v_{t(i-1)}, v_{ti}, v_{t(i+1)}, \dots, v_{t(i+k)}\}, S_{ti} \in \mathbb{R}^{2k+1}$$

O número de voxels considerados acima e abaixo do voxel central v_{ti} varia segundo as características do volume sísmico sendo considerado.

Durante o processo de criação do conjunto de amostras, para cada voxel v_{ti} do volume é criada uma amostra s_{ti} correspondente, exceto para os k primeiros e últimos voxels de cada traço, voxels de borda onde essa vizinhança não está disponível.

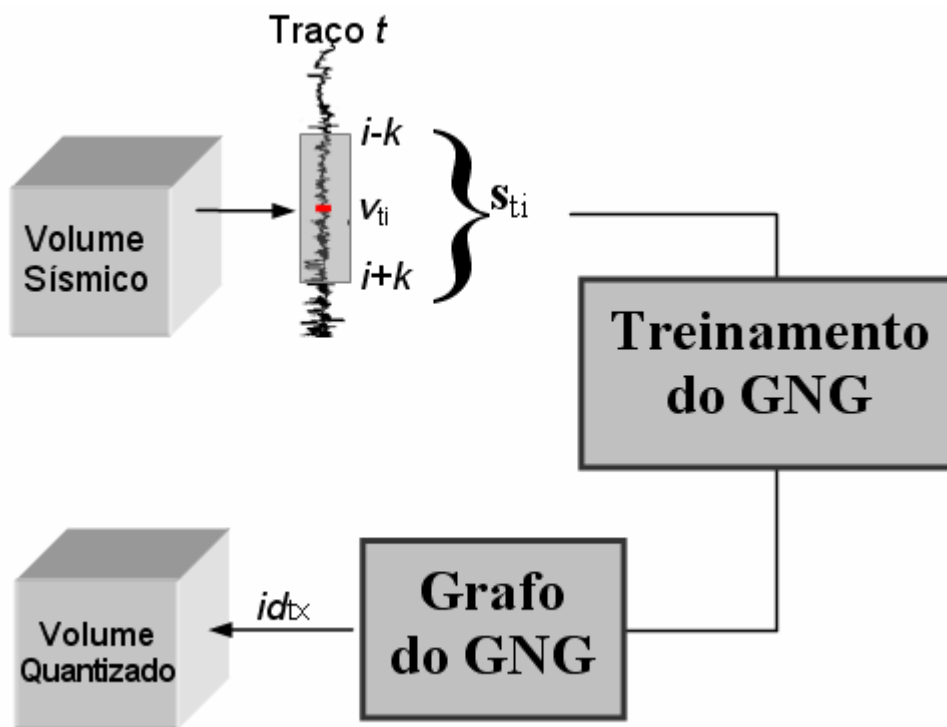


Figura 4.1 – A amostra s_{ti} is é adicionada ao conjunto de amostras utilizado para criar o grafo do algoritmo GNG. Os neurônios (nós) do grafo são utilizados para a criação do volume quantizado.

Depois de criadas todas as amostras, o conjunto de dados de entrada está completo. Esse conjunto de amostras é utilizado no processo de treinamento de uma instância do algoritmo GNG. Espera-se que a localização dos neurônios do grafo final obtido no espaço das amostras de entrada reflita fortemente a distribuição das amostras ao longo desse espaço.

4.2 Funções de similaridade adotadas

Conforme descrito anteriormente, uma característica que define diretamente a qualidade dos resultados obtidos pelo processo de agrupamento é o critério de medida que utilizamos para definir o quanto duas amostras são similares. Para dados sísmicos é necessário utilizar uma função que traduza adequadamente a proximidade existente entre duas amostras produzidas a partir do volume sísmico desejado. Duas funções de similaridade foram testadas. A primeira foi a distância

euclidiana, o critério de medida adotado no artigo original do GNG. Além dessa, também foi utilizado nos testes o coeficiente de correlação.

O coeficiente de correlação (r) entre duas amostras assume somente valores situados entre -1 e 1. O valor de r igual a 1 traduz a existência de uma correlação positiva entre os dois vetores. Da mesma forma, caso r seja -1 tem-se uma correlação negativa perfeita. Sejam duas amostras, \mathbf{s}_{t1a} e \mathbf{s}_{t2b} , respectivamente as amostras representando o a -ésimo voxel do traço $t1$ e o b -ésimo voxel do traço $t2$. O cálculo do coeficiente de correlação entre essas amostras é obtido segundo a seguinte fórmula:

$$r = \frac{\sum_{j=-k}^{j=+k} (v_{t1(a+j)} - m(s_{t1a})) \cdot (v_{t2(b+j)} - m(s_{t2b}))}{\sqrt{\sum_{j=-k}^{j=+k} (v_{t1(a+j)} - m(s_{t1a}))^2} \cdot \sqrt{\sum_{j=-k}^{j=+k} (v_{t2(b+j)} - m(s_{t2b}))^2}}$$

onde $m(\mathbf{s}_{t1a})$ e $m(\mathbf{s}_{t2b})$ são, respectivamente, os valores médios das amplitudes dos voxels das amostras \mathbf{s}_{t1a} e \mathbf{s}_{t2b} . A variável v_{t1} faz referência aos voxels da amostra \mathbf{s}_{t1a} e v_{t2} faz referência aos voxels da amostra \mathbf{s}_{t2b} .

4.3 A geração do novo volume quantizado

O algoritmo de treinamento do GNG será executado até que o número p desejado de nós seja atingido. Uma vez finalizado o processo de treinamento, os nós (neurônios) do grafo gerado resultam numa lista de p vetores de código correspondentes, e recebem um identificador numérico inteiro entre 1 e p , sendo utilizados para criar uma versão quantizada do volume de entrada. Nesse novo volume cada voxel v_{ii} é substituído pelo vetor de código \mathbf{c}_x do nó \mathbf{n}_x do grafo, onde sua amostra correspondente \mathbf{s}_{ii} foi classificada. No novo volume, cada voxel possui um valor numérico entre 1 e p .

4.4 Uma interpretação geométrica do volume quantizado

Como pode ser visto na figura 4.2, voxels de um mesmo horizonte, por terem características parecidas entre si, são representados por vetores de amostras que tendem a ser localizados em regiões similares do espaço \mathcal{R}^{2k+1} .

Na figura 4.2(a) vemos uma representação de sete vetores de amostras, todas representando voxels de um mesmo horizonte (em azul). Na figura 4.2(b) vemos sete curvas correspondendo às amplitudes dessas amostras. Todas essas 7 amplitudes foram classificadas em um mesmo nó, e na figura 4.2(c) é mostrado o vetor de código desse nó. Apesar de, pelas características do algoritmo de agrupamento, ser possível obter amostras referentes a voxels de um mesmo horizonte representadas por nós diferentes, essas amostras tendem a ser representadas por nós cujos vetores de código são próximos entre si.

4.4.1 Erro médio dos neurônios

A partir do grafo gerado, cada neurônio \mathbf{n}_x passa a representar seu conjunto de amostras, \mathbf{S}_x , todas amostras no espaço \mathcal{R}^{2k+1} mais próximas desse neurônio do que de qualquer outro. Cada uma dessas amostras passa a ser representada pelo vetor de código do neurônio \mathbf{n}_x , \mathbf{c}_x . Dessa forma, \mathbf{c}_x representa um conjunto \mathbf{S}_x de j amostras. Esse processo de representação gera um erro.

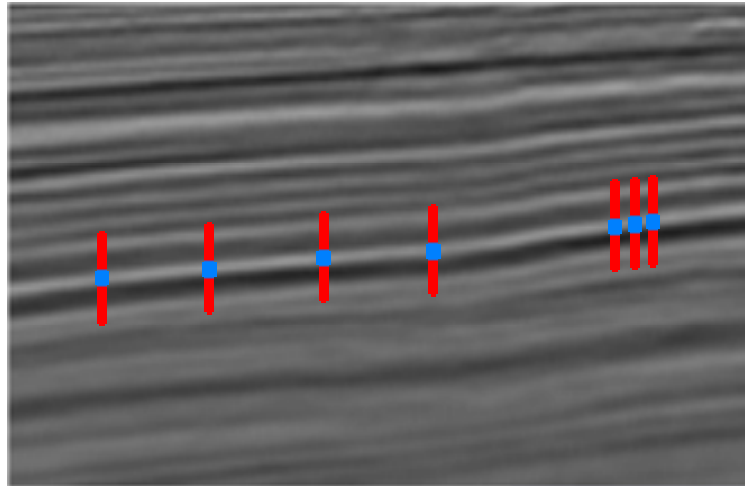
O erro total de \mathbf{n}_x com respeito ao seu conjunto \mathbf{S}_x de amostras poderia ser computado simplesmente como a soma das distâncias entre cada uma das amostras de \mathbf{S}_x e \mathbf{c}_x . Contudo, pelo fato de o número de amostras representadas pelos neurônios ser bastante variável, um critério mais adequado pode ser computar o erro médio (distância média) de todas as amostras de \mathbf{S}_x em relação a \mathbf{c}_x . Ou seja:

$$err_x = \frac{\sum_{i=0}^{j-1} dist(s_{xi}, c_x)}{j}$$

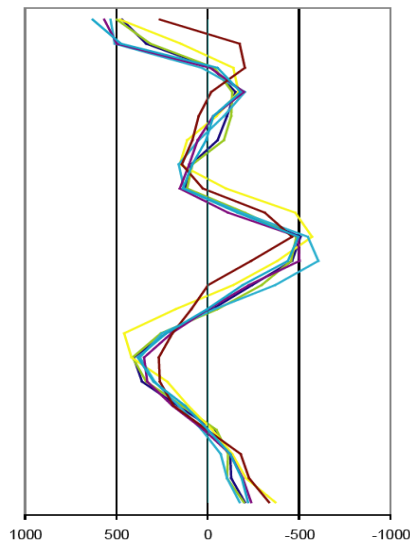
onde s_{xi} representa a i -ésima amostra do conjunto de amostra S_x do neurônio n_x , e $dist$ representa a função de similaridade (ou distância) sendo utilizada.

Depois de executado o processo de criação do volume resultante da quantização, é desejável que esse volume possa ser visualizado. No entanto, no novo volume, o conjunto de amplitudes dos voxels foi substituído por um conjunto de identificadores numéricos. Dessa forma, é necessário estabelecer uma escala de cores, com uma cor específica correspondente a cada identificador, ou seja, atribuir cores ao vetor de código de cada um dos neurônios da instância de GNG. Simplesmente atribuir aleatoriamente uma cor a cada um dos neurônios não é uma boa estratégia, uma vez que dessa forma as cores do padrão adotado não refletem a possível similaridade entre neurônios.

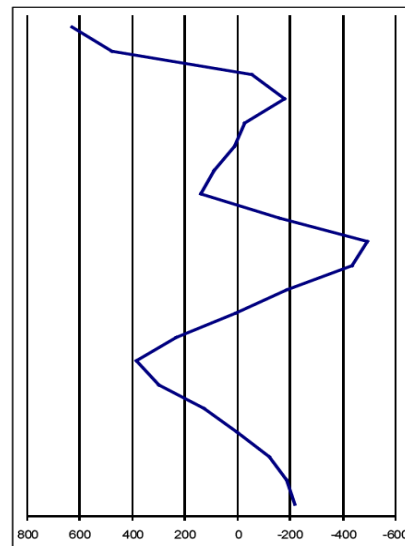
Uma boa forma de estabelecer uma escala de cores é atribuir a cor do neurônio segundo seu erro médio em relação às amostras que ele representa. Dessa forma, neurônios vizinhos (espera-se que vizinhos topológicos estejam definidos em áreas do espaço amostral onde a densidade das amostras é parecida) tenderão a ser representados por cores próximas. A figura 4.3 mostra um exemplo da visualização resultante desse processo. Na parte superior da figura, podemos ver uma fatia extraída do volume de entrada. Na parte inferior vemos a mesma fatia, obtida após o processo classificação das amostras e criação do volume quantizado. Os horizontes, bem como as falhas sísmicas, ficam mais bem definidos e pequenas ambigüidades existentes no dado original são minimizadas de maneira a diminuir seu impacto.



(a) Amostras de um mesmo horizonte sísmico



(b) Amplitudes das amostras



(c) Vetor de código correspondente

Figura 4.2 – Amostras de um mesmo horizonte e o vetor de código correspondente.

4.5 O método de extração de horizontes sísmicos

O algoritmo GNG é definido como um método não-supervisionado, e portanto não existe um controle explícito do processo de treinamento. Como resultado, o processo de quantização pode classificar amostras correspondentes a voxels de horizontes diferentes em um mesmo neurônio, especialmente se durante o processo de treinamento é escolhido um número reduzido de neurônios em

relação ao número de horizontes existentes no dado sísmico. Dois voxels não contíguos de um mesmo traço que possuam amplitudes e vizinhanças verticais semelhantes também podem ter o mesmo neurônio como representante. Por isso, o processo de mapeamento dos horizontes não pode ser baseado em somente um voxel do volume quantizado.

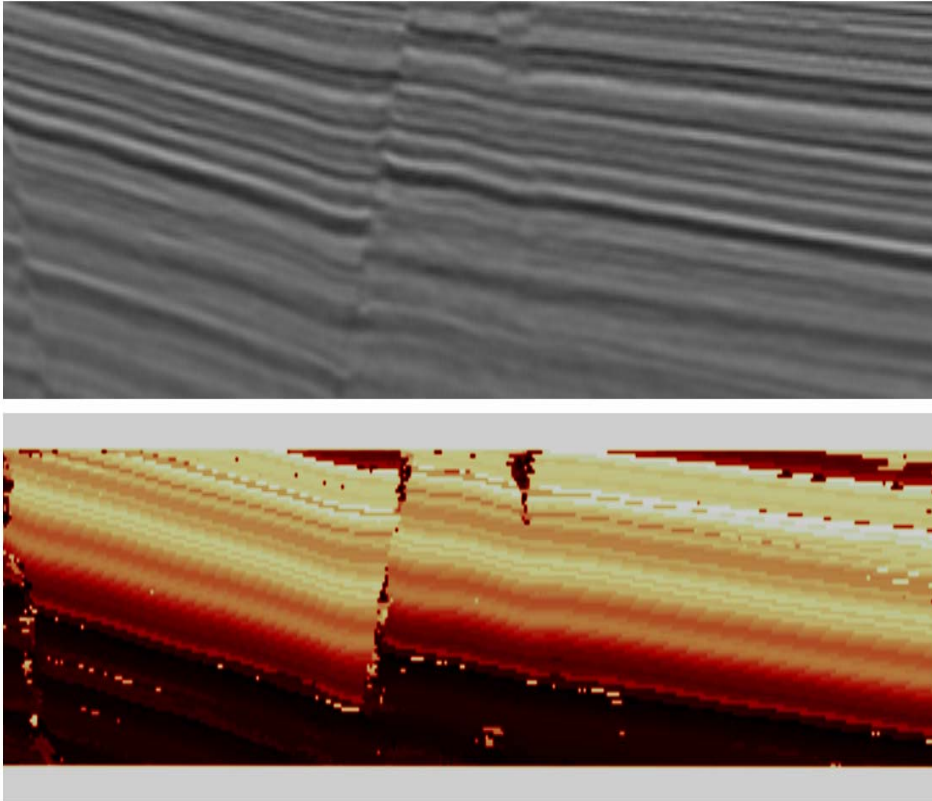


Figura 4.3 – Horizontes e falhas sísmicas no volume quantizado. Ilustração da mesma fatia no volume de entrada e no volume quantizado.

Para suplantarmos essa limitação, propomos que cada voxel v_{ti} , o i -ésimo voxel do traço t do volume quantizado, seja representado por uma amostra \mathbf{f}_{ti} , formada de maneira similar à criação das amostras \mathbf{s}_{ti} , com a diferença de que agora os componentes são obtidos a partir do volume quantizado. Podemos definir essas novas amostras, vetores de inteiros, como:

$$\mathbf{f}_{ti} = \left(id_{x(i-q)}, \dots, id_{xi}, \dots, id_{x(i+q)} \right)^T \in \mathbf{Z}^{2q+1}$$

onde q é o número de identificadores de neurônios acima e abaixo de v_{ti} .

Para encontrar os voxels pertencentes ao horizonte a ser mapeado, uma nova função de similaridade é definida no espaço das novas amostras. O processo de extração inicialmente recebe um voxel v_{ta} (uma semente) do horizonte que se deseja mapear e sua amostra correspondente \mathbf{f}_{ta} , ambas definidas a partir de um traço t qualquer do volume. O objetivo do procedimento é encontrar amostras similares a \mathbf{f}_{ta} , denominadas aqui genericamente como \mathbf{f}_b , ao longo dos demais traços. Para isso, a função de similaridade utiliza uma função D que recebe \mathbf{f}_{ta} e \mathbf{f}_b .

Essa função D poderia ser modelada de diversas formas. Uma primeira forma D_1 seria, uma vez recebidas \mathbf{f}_{ta} e \mathbf{f}_b , simplesmente somar as distâncias dos vetores de código correspondentes aos identificadores inteiros de cada uma de suas componentes. No entanto, essa abordagem não leva em conta a densidade amostral local do conjunto de amostras de entrada \mathbf{S} .

A localização dos nós do grafo do GNG reflete a densidade local das amostras de entrada no espaço de entrada inicial. Isso significa que neurônios localizados no espaço \mathcal{R}^{2k+1} em regiões de maior concentração de amostras por unidade de volume são relativamente mais próximos dos seus vizinhos do que aqueles situados em regiões do espaço onde a densidade amostral é mais baixa. Portanto, para comparar distâncias entre neurônios precisamos dar maior ou menor peso de acordo com sua posição espacial.

O grafo formado pelo algoritmo GNG inclui uma lista de arestas ligando neurônios próximos (vizinhos topológicos). Caso uma aresta ligue dois vizinhos topológicos (independentemente da densidade amostral dessa região), a função D utiliza a média d_{mx} das distâncias existentes entre um neurônio \mathbf{n}_x e seus vizinhos topológicos. A média d_{mx} é definida simplesmente como:

$$d_{mx} = \frac{1}{m_x} \sum_{i=0}^{m_x-1} dist(\mathbf{c}_i, \mathbf{c}_x),$$

onde m_x é o número de vizinhos topológicos do neurônio \mathbf{n}_x e \mathbf{c}_i representa o vetor

de código do i -ésimo vizinho topológico de \mathbf{n}_x . Uma vez calculada essa média para todos os neurônios do grafo, a distância do neurônio \mathbf{n}_x em relação a um segundo neurônio \mathbf{n}_y pode ser dada por:

$$d_{xy} = \frac{1}{d_{mx}} \text{dist}(\mathbf{c}_y, \mathbf{c}_x)$$

A função D entre \mathbf{f}_{ta} e \mathbf{f}_b utilizada pelo procedimento de extração leva em conta ainda outra particularidade dos dados sísmicos, e não somente a soma das distâncias d_{xy} dos vetores de código correspondentes. Para cada componente j de \mathbf{f}_{ta} é permitida a comparação entre os componentes $j-1$, j , $j+1$ de \mathbf{f}_b , e somente a menor das três distâncias é utilizada na soma dos componentes. Essa tolerância pode ser maior em volumes sísmicos onde as camadas possuam inclinações maiores. Na figura 4.4 tem-se uma ilustração do procedimento de tolerância. Na figura o j -ésimo componente (j -ésimo vetor de código) de \mathbf{f}_{ta} , id_{aj} , é comparado com as amostras $id_{b(j-tol)}$, id_{bj} e $id_{b(j+tol)}$ da amostra \mathbf{f}_b . A menor distância, que no caso da ilustração utiliza o identificador $id_{b(j+tol)}$, é a distância utilizada em D .

O procedimento D é executado para todos os traços exceto o próprio traço que contém a semente \mathbf{f}_{ta} , e para cada um dos traços a amostra mais próxima \mathbf{f}_b é armazenada, passando a compor o conjunto \mathbf{F} de voxels candidatos.

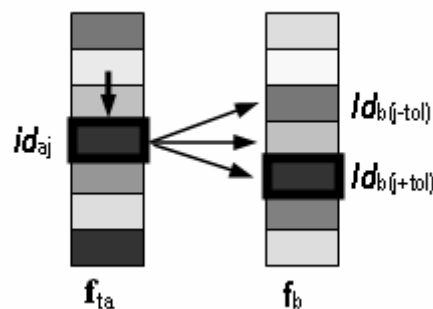


Figura 4.4 – O procedimento D de similaridade entre duas amostras. Tal procedimento recebe duas amostras, \mathbf{f}_{ta} e \mathbf{f}_b .

Depois de formado, o conjunto \mathbf{F} possui o melhor candidato de cada um dos traços do volume. A z -ésima amostra de \mathbf{F} , \mathbf{f}_z , representa o seu voxel v_z correspondente e somente é considerado parte do horizonte definido a partir da semente correspondente a v_{ta} caso a similaridade entre \mathbf{f}_{ta} e \mathbf{f}_z , D_{az} , seja menor que um valor de tolerância T_1 e todos os outros voxels do mesmo traço (exceto vizinhos imediatos) resultem em distâncias maiores que outro valor de tolerância T_2 .

Com a adoção desse procedimento, desde que as tolerâncias e os comprimentos das amostras sejam bem dimensionados, espera-se que os voxels retornados realmente pertençam ao conjunto de voxels que formam o horizonte sendo mapeado. No entanto, durante o processo de desenvolvimento do algoritmo de extração, foi constatado que, apesar de o conjunto de voxels obtido pelo algoritmo representar satisfatoriamente o horizonte, voxels situados aleatoriamente dentro das regiões de falha presentes no volume eram esporadicamente retornados. Tal possibilidade é explicada pela natureza caótica das falhas sísmicas. Esses falsos positivos, no entanto, tendem a ser completamente eliminados caso sejam comparados com seus vizinhos laterais. Para isso, depois de encontrar cada um dos voxels do horizonte, é encontrada a distância entre esse voxel e seus vizinhos laterais situados nos traços vizinhos. Caso essa distância seja maior que um valor de tolerância T_3 , esse voxel é descartado.

4.6 Mapeamento de falhas sísmicas

As falhas sísmicas são quebras na continuidade original dos horizontes, causando um deslocamento relativo das rochas e fazendo com que elas percam sua continuidade original. Na figura 4.5 apresentamos um exemplo de uma falha desse tipo. Na parte superior da figura mostramos uma fatia do volume onde a falha não está presente. Já na parte inferior mostramos uma outra fatia do mesmo volume sísmico onde uma falha sísmica pode ser verificada. Os retângulos vermelhos demarcam a região onde a presença da falha é constatada.

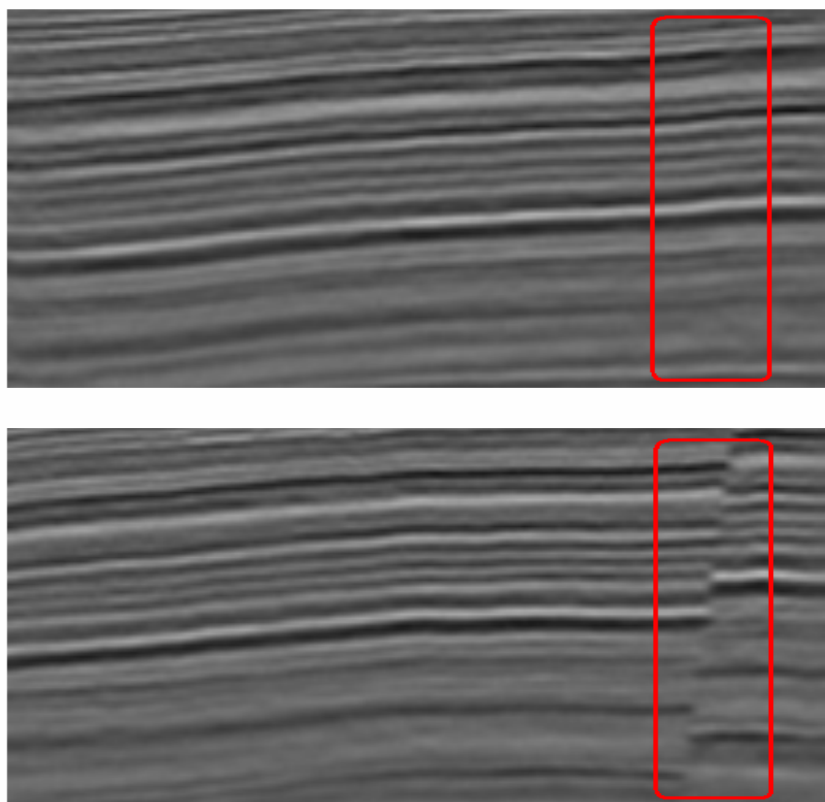


Figura 4.5 – Duas fatias de um mesmo volume sísmico. Na segunda uma falha sísmica é evidenciada (retângulo vermelho).

Nos dados sísmicos, os voxels de um horizonte podem ser reconhecidos por sua alta coerência espacial em relação aos seus vizinhos. Com relação às falhas sísmicas, a quebra de continuidade é acompanhada de fenômenos como compressão e erosão, aumentando a natureza caótica da região onde elas ocorrem e fazendo com que essas regiões, ao contrário das regiões onde existem horizontes definidos, sejam caracterizadas por um alto nível de distorção nos dados. Por isso os voxels das regiões de falha são caracterizados pela sua pouca coerência espacial em relação aos seus voxels vizinhos. Uma vez que a criação do volume quantizado tende a diminuir o nível de ruído presente nos dados originais, nesses volumes as características de coerência dos voxels de ambas as estruturas geológicas (tanto horizontes quanto falhas) tendem a ser fortemente evidenciadas. Isso pode ser verificado na figura 4.6, onde vemos duas fatias ortogonais cortando um volume quantizado. Fica evidenciada a coerência dos voxels vizinhos dos horizontes e a quebra da continuidade das superfícies nas regiões onde existe a presença de falhas sísmicas.

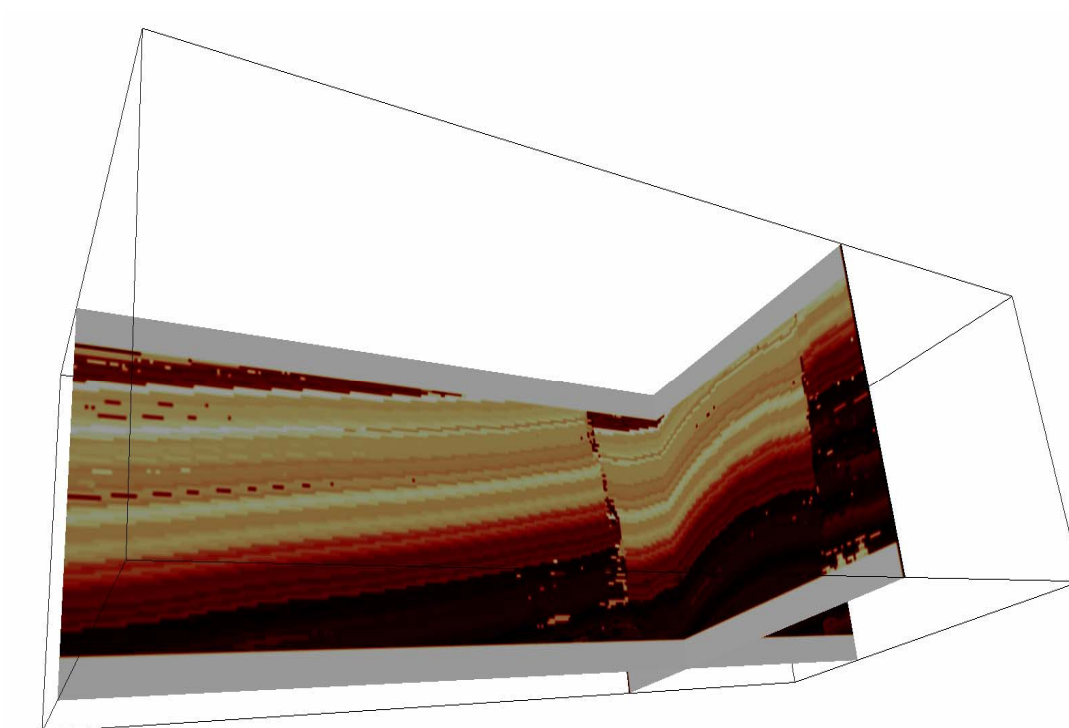


Figura 4.6 – Duas fatias ortogonais demonstrando a coerência espacial entre voxels vizinhos, exceto por aqueles que pertencem às falhas sísmicas.

Essa boa qualidade da classificação dos voxels obtida pela criação do volume quantizado serviu de motivação para o desenvolvimento do algoritmo de extração de falhas aqui descrito. Tal procedimento mostrou resultados iniciais bastante promissores. O algoritmo se inicia a partir de um voxel semente recebido como parâmetro, e utiliza como critério de similaridade o mesmo procedimento D definido para o processo de mapeamento dos horizontes.

Para verificar se um determinado voxel v_{ti} pertence a uma falha, primeiramente é necessário definir as dimensões de um paralelepípedo ao redor desse voxel, que envolve todos os vizinhos de v_{ti} que serão considerados no cálculo. Os vizinhos situados no mesmo traço sísmico do voxel candidato não são levados em conta. Na figura 4.7 pode-se ver um exemplo de conjunto de voxels ao redor do voxel v_{ti} candidato a falha (em vermelho) onde somente os vizinhos imediatos (em cinza) estão sendo considerados.

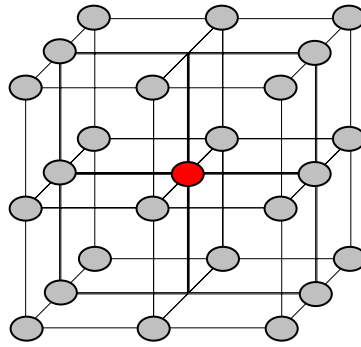


Figura 4.7 – O voxel candidato a pertencer a uma falha (vermelho) e seus vizinhos imediatos (em cinza).

Definidos os vizinhos, é calculado o valor da similaridade D da amostra \mathbf{f}_i que representa o voxel v_{ti} em relação às amostras que representam seus voxels vizinhos.

Depois de calculado esse conjunto de valores de similaridade, vários critérios podem ser adotados para determinar se v_{ti} pertence ou não a uma falha sísmica. Um exemplo possível seria encontrar a média dos valores de similaridade obtidos. Um voxel pouco similar a seus vizinhos próximos tende a estar em uma região de pouca coerência espacial, o que caracterizaria a falha. Segundo esse critério, o voxel seria classificado como pertencente à falha caso a média obtida fosse maior que um valor pré-estabelecido.

Caso um voxel pertença a um horizonte qualquer do volume, é esperado que na sua lista de vizinhos laterais existam voxels pertencentes a esse mesmo horizonte. Levando isso em conta, no procedimento que adotamos encontra-se, a partir da lista de voxels vizinhos, aquele voxel que é o mais similar a v_{ti} . Caso mesmo esse vizinho possua um valor de similaridade maior que um valor de tolerância T_m estabelecido, considera-se que naquele ponto não existe um horizonte definido, e v_{ti} é considerado um voxel de falha.

O procedimento se inicia a partir do voxel semente recebido. Caso esse voxel seja considerado falha, é imediatamente adicionado à lista de voxels que representam a falha em questão, e seus vizinhos são incluídos na lista de voxels candidatos que serão testados posteriormente. O processo é repetido para cada um

dos voxels da lista de candidatos, até que essa lista esteja vazia.

Ao final, tem-se um conjunto de voxels formando uma nuvem de pontos conexa onde a falha mapeada está definida, e espera-se que essa nuvem de pontos defina a região de falha com precisão. Tal conjunto de pontos, apesar da boa precisão obtida nos testes, não forma uma superfície simples, diferindo do resultado obtido pelo procedimento de mapeamento manual das falhas, que geralmente irá produzir somente uma superfície simples de falha. O processo de obtenção de uma superfície a partir da nuvem de voxels de falha obtida não foi desenvolvido durante a pesquisa para esta dissertação.