# 3D pyramid interpolation

*Xukai Shen*

## ABSTRACT

Seismic data geometries are not always as nice and regular as we want due to various acquisition constraints. In such cases, data interpolation becomes necessary. Usually high-frequency data are aliased, while low-frequency data are not, so information in low frequencies can help us interpolate aliased high-frequency data. In this paper, I present a 3D data interpolation scheme in pyramid domain, in which I use information in low-frequency data to interpolate aliased high-frequency data. This is possible since in pyramid domain, only one prediction error filter (PEF) is needed to represent any stationary event (plane-wave) across all offsets and frequencies. However, if we need to estimate both the missing data and PEF, the problem becomes nonlinear. By alternately estimating the missing data and PEF, we can linearize the problem and solve it using a conventional least-squares solver.

## INTRODUCTION

Data interpolation is an important step in seismic data processing that can greatly affect the results of later processing steps, such as multiple removal, migration and inversion. There are many ways to interpolate data, including Fourier-transform-based approaches (e.g., Xu et al., 2005) and PEF-based approaches (e.g., Spitz, 1991; Crawley, 2000). A PEF is a filter that predicts one data sample from $n$ previous samples, where $n$ is the length of the PEF. One important feature of a PEF is that it has the inverse spectrum of the known data, so when it is convolved with known data, it minimizes the convolution result in the least-square sense. PEF estimation can be done in either time-space ($t$-$\mathbf{x}$) domain or frequency-space ($f$-$\mathbf{x}$) domain (e.g. Claerbout, 1999; Crawley, 2000; Curry, 2007), however, if PEF estimation is done in the $f$-$\mathbf{x}$ domain, every frequency needs one distinct PEF.

The pyramid domain was introduced by Ronen (Hung et al., 2005), and is a resampled representation of an ordinary $f$-$\mathbf{x}$ domain. Although it has frequency and space axes, the spatial sampling is different for different frequencies. This is attractive because we can use sparser sampling to adequately sample the data at lower frequencies, which makes uniform sampling for all frequencies unnecessary. Therefore in the pyramid domain, coarser grid spacing is used for lower frequencies, while finer spacing is used for higher frequencies. This makes it possible to capture the character of all frequency components of stationary events with only one PEF. So the information in the low frequency data can be better used to interpolate higher frequency data.

In this paper, I present a 3D version of data interpolation in the pyramid domain based on PEF estimation, which is based on Shen (2008). The paper is organized as follows: I first show the 3D pyramid transform and corresponding missing-data interpolation and PEF estimation. I then show synthetic data examples. Finally, I conclude with the advantages and disadvantages of this interpolation method.

# METHODOLOGY

There are two important parts of the pyramid-based interpolation algorithm, the first of which is the selection of the pyramid transforms between pyramid domain and $f$-$\mathbf{x}$ domain. The more accurately these transforms are performed, the better the result we can get for missing data interpolation. The second step combines data interpolation and PEF estimation in the pyramid domain, which are done alternately in an iterative way.

## 3D pyramid Transform between pyramid domain and $f$-x domain

I discussed the 2D pyramid transform in Shen (2008), and the 3D version is almost the same, except that some scalars become vectors. In the 3D pyramid transform, spatial grid spacing is calculated for each frequency $f$ using the equation

$$\Delta \mathbf{x}(f) = \frac{\Delta \mathbf{x_0} \mathbf{v}}{f \mathbf{n_{sf}}}, \tag{1}$$

where $\Delta \mathbf{x}(f)$, $\Delta \mathbf{x_0}$, $\mathbf{v}$ and $\mathbf{n_{sf}}$ are all 2D vectors. $\Delta \mathbf{x_0}$ is the uniform spatial grid spacing in the original $f$-$\mathbf{x}$ data, $\mathbf{v}$ is the velocity that controls the slope of the pyramid and $\mathbf{n_{sf}}$ is the sampling factor in pyramid domain. By changing this factor we can control how densely the pyramid domain is sampled. In situations where events to be interpolated are not perfectly stationary, dense sampling is preferrable since the information in the low frequencies cannot be represented well by only a few points. In 3D, the inversion scheme that transforms data in $f$-$\mathbf{x}$ space to the pyramid domain is as follows:

$$\mathbf{Lm} - \mathbf{d} \approx \mathbf{0}, \tag{2}$$

where $\mathbf{m}$ is the data in the pyramid domain, $\mathbf{d}$ is the known data in $f$-$\mathbf{x}$ space, and $\mathbf{L}$ is the 2D linear interpolation operator in 3D pyramid transform. The 3D pyramid transform from pyramid domain to $f$-$\mathbf{x}$ domina uses the following equation :

$$\mathbf{d} = \mathbf{Lm}. \tag{3}$$

Where now $\mathbf{m}$ is known and $\mathbf{d}$ is unknown data in $f$-$\mathbf{x}$ space.

## PEF estimation and Missing data estimation

The missing data estimation algorithm presented here is different from what I presented in the previous paper (Shen, 2008). Missing data are fitted in the $f$-$\mathbf{x}$ domain to ensure better fitting of known data. Also, the 3D version of these algorithms use helical coordinates (Claerbout, 1999) to perform the convolution.

For PEF estimation, I try to solve the following problem assuming known pyramid data $\mathbf{m}$. Denoting convolution with m as operator $\mathbf{M}$, with $\mathbf{W}$ being a diagonal masking matrix that is 1 where pyramid data can be used for PEF estimation and 0 elsewhere, I try to solve for the unknown PEF $\mathbf{a}$ using the following fitting goal(Claerbout, 1999):

$$\mathbf{WMa} \approx \mathbf{0}, \tag{4}$$

For missing-data estimation, I start with a known PEF $\mathbf{A}$, and try to solve the following least-squares problem (Claerbout, 1999):

$$\begin{aligned} &\mathbf{K}\left(\mathbf{Lm}-\mathbf{d}\right) \approx \mathbf{0} \\ &\epsilon\mathbf{WAm} \approx \mathbf{0}, \end{aligned} \tag{5}$$

where $\mathbf{K}$ is a diagonal masking matrix that is 1 where data is known and 0 elsewhere, $\epsilon$ is a weight coefficient that reflects our confidence in the PEF, and $\mathbf{W}$ is the same as explained above.

## Linearized nonlinear problem

To estimate both missing data and the PEF, the problem becomes nonlinear. To avoid directly solving nonlinear problem, I linearize it by alternately estimate missing data and PEF, and use them to update each other. Corresponding pseudo code is as follows:

```
for each iteration i{
 estimate missing data m_i from PEF a_(i-1) using equation 5
 estimate PEF a_i from missing data m_i using equation 4
}
```

I start with my guess of the PEF, $\mathbf{a_0}$. First, I make an operator $\mathbf{A_0}$ that is convolution with $\mathbf{a_0}$, and I use it to estimate the missing data $\mathbf{m_0}$. From $\mathbf{m_0}$, I make an operator $\mathbf{M_0}$ that is a convolution with $\mathbf{m_0}$. Then I update $\mathbf{a_0}$ using $\mathbf{M_0}$, calling the updated $\mathbf{a_0}$ as $\mathbf{a_1}$. This process makes one iteration of the linearized problem. Then I repeat this process, making $\mathbf{A_1}$ from $\mathbf{a_1}$, updating $\mathbf{m_0}$ to $\mathbf{m_1}$ using $\mathbf{A_1}$, making $\mathbf{M_1}$ from $\mathbf{m_1}$, updating $\mathbf{a_1}$ to $\mathbf{a_2}$ using $\mathbf{M_1}$, and so on... Finally the algorithm will converge to some $\mathbf{m}$ and $\mathbf{a}$; hopefully, by careful choosing of $\mathbf{a_0}$, I will converge to the correct $\mathbf{m}$ and $\mathbf{a}$.

# EXAMPLE

Here I show two examples of missing-data interpolation using the linearized iterations described above. Both are 3D synthetic examples. The first example interpolates 3D plane-waves. The second dataset is a patch from the qdome data set Claerbout (1999).

## Synthetic plane-waves

In this example, there are three plane-waves with different frequency components and dips (Figure 1a). I sub-sampled the data cube by a factor of three along both the $x$ and $y$ axes. This causes aliasing in two of the three plane-waves (Figure 1b), which can also be seen from their $f$-$k$ spectrum( Figure 2). Then with the initial guess of the PEF being a 2D Laplacian operator, the above algorithms converged to a decent result at most places, except for data points close to the edges (Figure 3a). More specifically, the remaining data, after being transformed into the pyramid domain, looks like Figure 4a. There are a lot of holes, which are caused by missing data and big sampling factors ($\mathbf{n_{sf}}$ in equation 1) used in pyramid domain. In this example, a sampling factor of 6 along the inline direction and a factor of 8 along crossline direction are used to make sure we have enough sample to represent all the frequencies, especially low frequencies. With the initial guess of the PEF being a 2D Laplacian operator, these holes in the pyramid domain will be filled; however, since this is just a guess of the PEF, the filled information is not necessarily correct (Figure 5b). Actually, the $t$-$\mathbf{x}$ domain data of interpolation with this PEF are step-like functions for aliased plane-waves (Figure 3b). However, for the unaliased low-frequency plane-wave, the missing data is already correctly interpolated. After five iterations of the algorithm, using the information from low-frequency data, the interpolated data finally have the correct dips for all the frequency components (Figure 5a).

## Patch from the qdome data set

In this example, I windowed out one fourth of the qdome data set, in which the reflectors are almost stationary (Figure 6 a). I then sub-sampled the data cube by a factor of four along both the $x$ and $y$ axes (Figure 6 b). This caused aliasing of some reflectors, especially in the cross line direction, as the $f$-$\mathbf{k}$ spectrum of Figure 7 shows. Since this data set is more complicated then the previous one, and not all the reflectors are ideally stationary, in the pyramid domain, I use ten as the sampling factor along both $x$ and $y$ directions to make sure low frequency data are well represented. Then with the initial guess of the PEF also being 2D Laplacian operator, the above algorithms converged to a satisfactory result for most of the reflectors (Figure 8b ). Notice that for reflectors with small amplitude, the interpolation works not so well, and tuning the PEF size may help solving this problem. By looking at the the depth slice in Figure 8b , zeros can be observed for data at small $x$ and $y$ values, this is due to insufficient number of data points for PEF estimation at these

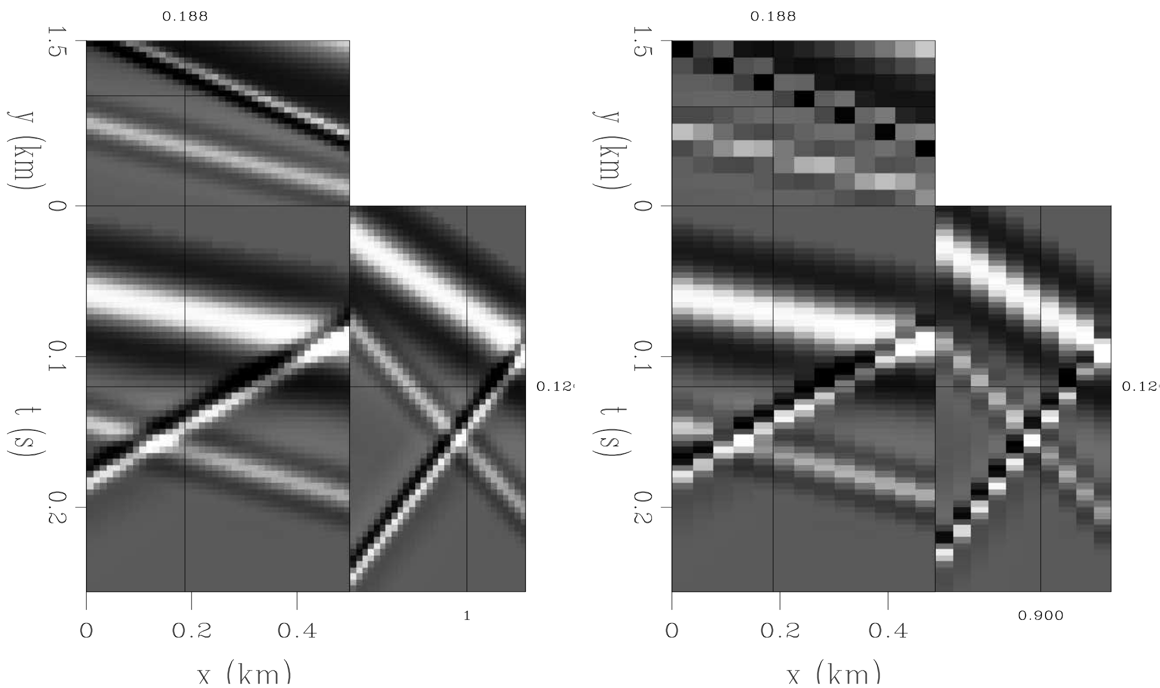


Figure 1: a) Original data, consisting of three plane waves with different dips and frequency contents. b) Sub-sampling by a factor of three along both $x$ and $y$ axes. [**ER**]

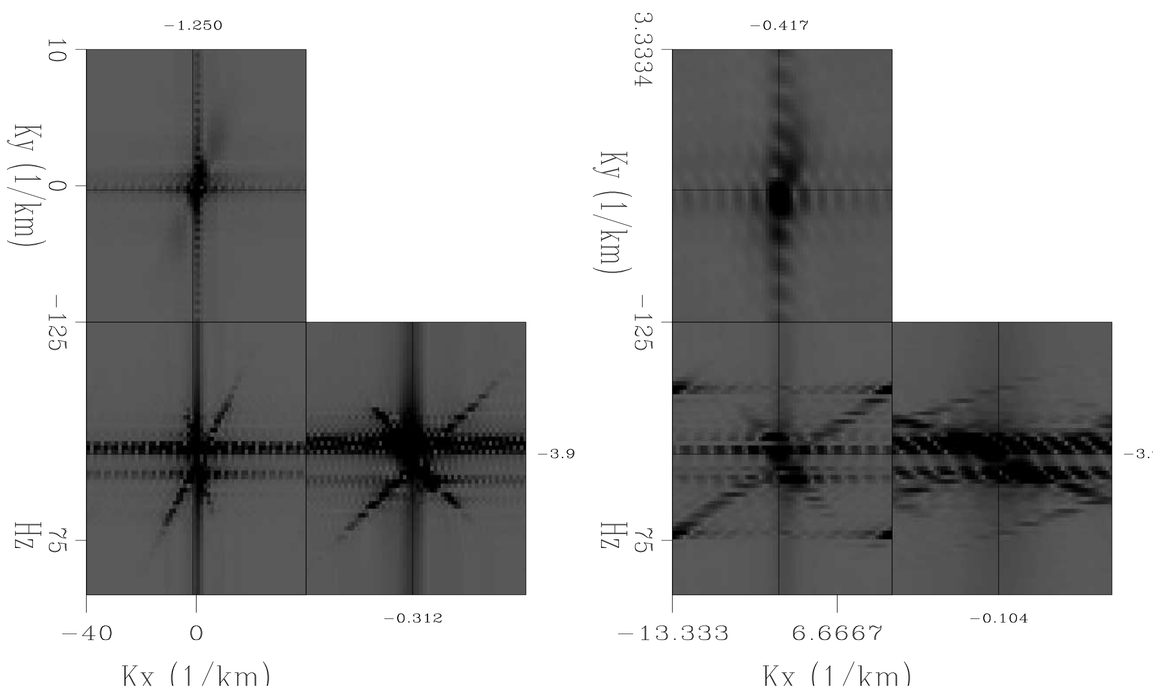


Figure 2: a) The $f$-$k$ spectrum of the original data. b) The $f$-$k$ spectrum of the sub-sampled data. [**ER**]

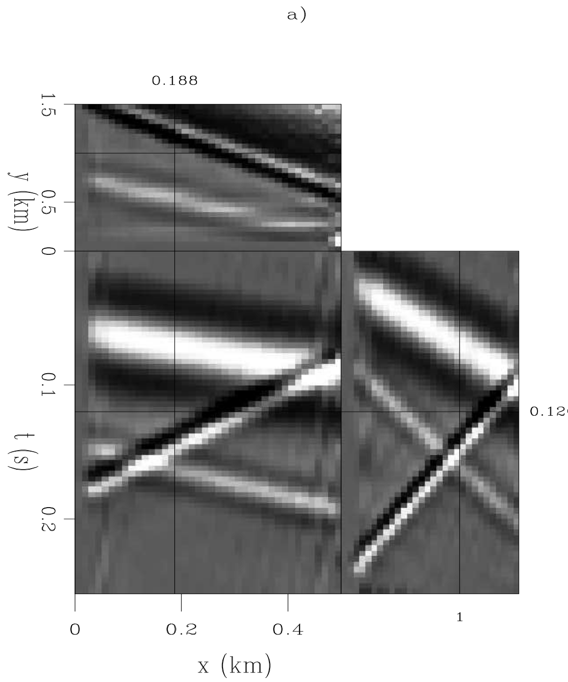


Figure 3: a) Interpolated data. b) Data interpolated with the Laplacian operator in the pyramid domain. [**CR**]

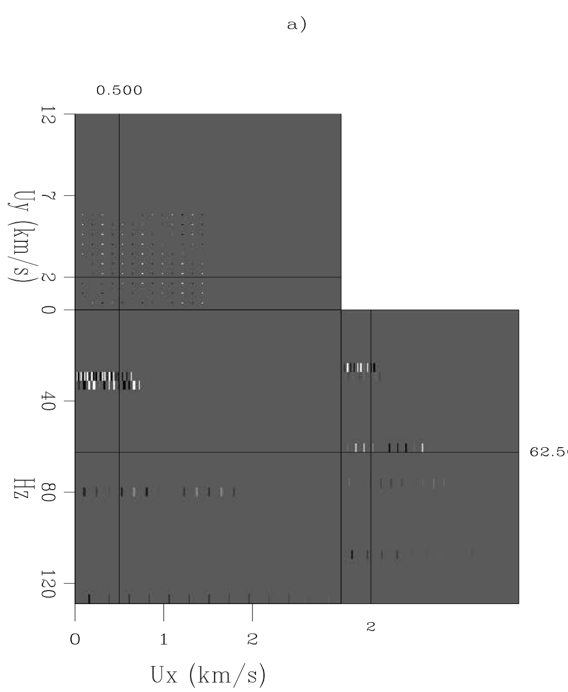


Figure 4: a) Remaining data in the pyramid domain. b) Data interpolated with the Laplacian operator in pyramid domain. [**CR**]

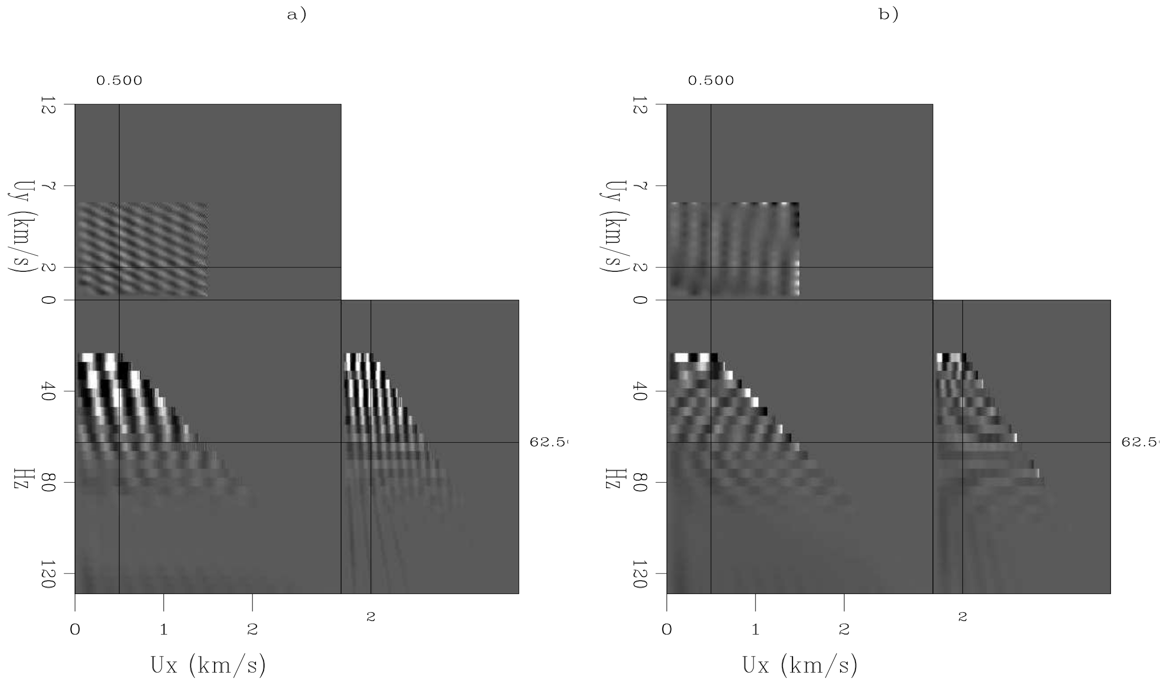


Figure 5: a) Interpolated data (final version) in pyramid domain. b) Data interpolated with the Laplacian operator in pyramid domain. [**CR**]

locations. On the other hand, the other two edges have much stronger amplitude artifacts, which requires further investigation.

## FUTURE WORKS

So far, for stationary events, this algorithm works quite well. For real data, the assumption of stationarity holds if we break the data into small patches and look at each individual patch. So the next step will be to looking at more complex data and try to apply this algorithm with patching technique.

## CONCLUSIONS

The pyramid domain is a very promising domain for missing-data interpolation. The synthetic examples demonstrate that, with a good initial PEF estimate, we can use the information in the low frequency to interpolate the aliased missing data relatively accurately and interpolate the unaliased missing data fairly well.

One disadvantage of this interpolation scheme is computation cost. First, to get a decent result, the data samples in pyramid domain is an order more than that in $f$-$\mathbf{x}$ domain along each spatial axis; in 3D, that amounts to a factor of 100 or more. In addition, the linearized nonlinear iteration adds a factor of about five in the synthetic

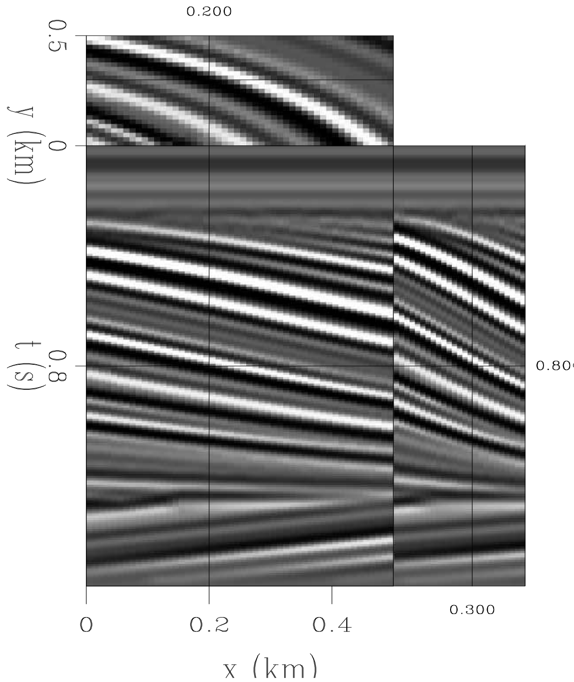


Figure 6: a) Original patch of qdome data, where all the reflectors are almost stationary. b) Sub-sampling by a factor of four along both the $x$ and $y$ axes. [**ER**]

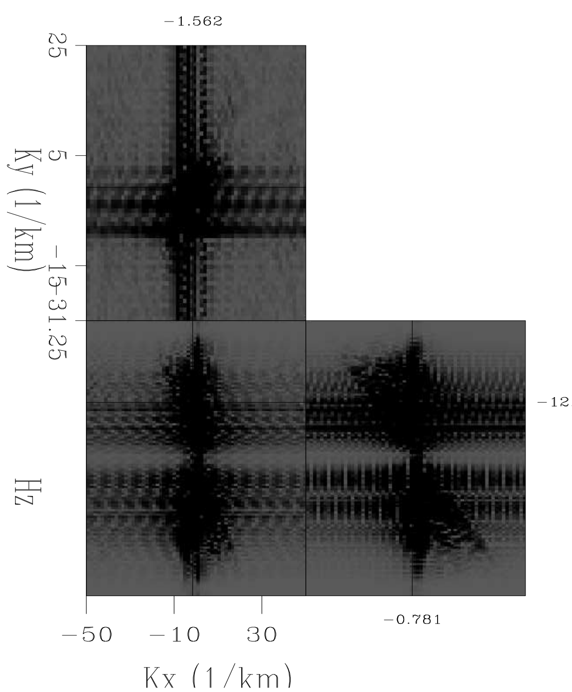


Figure 7: a) The $f$-$k$ spectrum of the original data. b) The $f$-$k$ spectrum of the sub-sampled data. [**ER**]

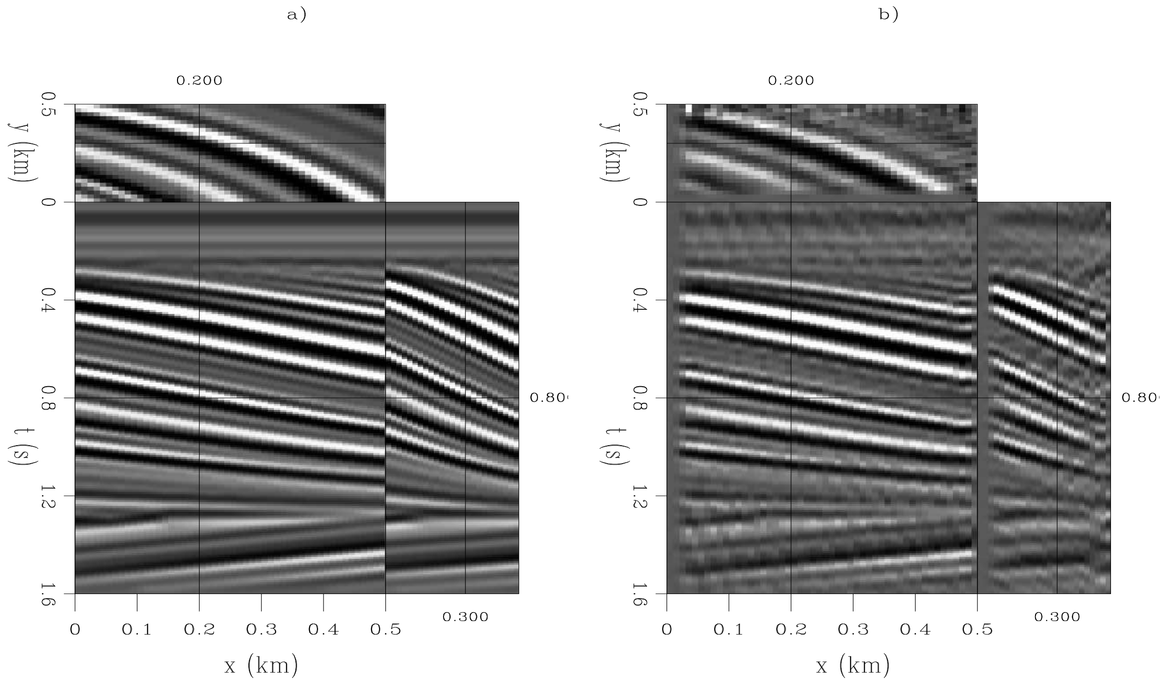


Figure 8: a) Original patch of qdome data. b) Interpolated data. [**CR**]

test. In other words, we have to do both PEF estimation and data interpolation five times in total. Altogether, we first increase the data size by a factor of 100, then run about 5 rounds of data estimation. So the overall computational cost is about 500 times greater than a conventional PEF based interpolation scheme (e.g. Spitz, 1991). However, with a patching technique, many patches of data can be interpolated simutaneously using parallelized version of this algorithm.

## ACKNOWLEDGMENTS

I would like to thank Jon Claerbout for ideas about data interpolation in the pyramid domain, Bob Clapp for synthetic example suggestions, and Antoine Guitton and William Curry for numerous debugging suggestions.

## REFERENCES

Claerbout, J., 1999, Geophysical estimation by example: Environmental soundings image enhancement: Stanford Exploration Project.

Crawley, S., 2000, Seismic trace interpolation with nonstationary prediction-error filters: PhD thesis, Stanford University.

Curry, W., 2007, f-x domain non-stationary interpolation in patches: SEP-Report, **131**.

Hung, B., C. Notfors, and S. Ronen, 2005, Seismic data reconstruction using the pyramid transform: 67th Meeting, EAGE, Expanded Abstracts, A039.

Shen, X., 2008, Data interpolation in pyramid domain: **SEP-134**, 225–230.

Spitz, S., 1991, Seismic trace interpolation in the F-X domain: Geophysics, **56**, 785–794.

Xu, S., Y. Zhang, D. Pham, and G. Lambare, 2005, Antileakage Fourier transform for seismic data regularization: Geophysics, **70**, V87–V95.