Accelerated Data Handling Using Real-Time Data Compression and Decompression for Data Sciences and Geoscience

Diderich Buch, Morten Vinther

Headwave™

79th EAGE Conference & Exhibition 2017

INTRODUCTION

Data complexity and size is growing exponentially and intelligent data handling is the only way to manage this. Current industry standard relies on raw data to be read from disc, handled/manipulated and then sent back to disc. This generates not only an enormous amount of data but also puts strain on network. Additionally, engineers spend valuable time waiting on data to load, process and save.

The VDS compressed data format significantly improves data handling, which makes data available 5 – 30x faster. The result is significant savings on both hardware cost and engineer time - purely from software.

MOTIVATION

The VDS compression retains the full fidelity (phase, frequency and amplitude) of the signal by exploiting correlation in wavelet space. For raw shot records, a special purpose integer based compression method is utilised.

The format has been in use for approximately a decade in commercial and R&D software. The VDS format is generic and can be used for:

2D, 3D (pre- and post stack), 4D and 5D (WAZ) seismic in addition to VSP. The compression works on standard hardware with no need for costly upgrades.

Data sciences experts advocate use of all available data. This is a huge challenge in itself. In addition, common 'big data' environments such as Hadoop generally require data to be duplicated or in some cases (e.g. Spark) loaded into distributed memory.

In this poster we will discuss examples that show that there are significantly smarter ways to consume all the available data, both within big data environments as well as with commonly used toolkits in data science. This is critical both for learning and inference, as well as for more general statistical methods.

VDS
A Modern Data Container

- On demand levels of decompression (lossless to lossy)
- "Instant" random access
- High performance I/O
- Full metadata (attribute mark-up, etc)
- Full E&P workflow: from processing to interpretation and beyond

METHOD - 3 STEP PROCESS

Tiling

- Configurable tile size
- Default 128x128x128, 2 voxel overlap
- Can also compress 2D-slices for inline/crossline/time slice

Wavelet Transform (Lossy)

- Standard CDF 9/7 transform
- Implemented by the lifting scheme
- Similar (better) approach as lossy mode of JPEG 2000

Quantization & Encoding or Decoding

- Does not clip data outside of value range
- Proprietary algorithm – higher quality compared to
- Set Partitioning Trees Wavelet encoding
- Very efficient parallel (OpenMP/CUDA) encode and decode

RESULTS / APPLICATION

Using the compression dramatically reduces both the performance bottlenecks in relation to processing/imaging input output and storage requirements. Also, importantly, it reduces any network congestion which again can improve remote office work and any cloud initiatives.

This accelerate the access to massive data such as WAZ seismic data that keeps increasing in size. Any size data can with the compression be accessed much faster and is not limited by CPU or GPU memory. Having such faster access to data speeds up end-user computations and visualization that enables an interactive quantitative problem solving routine.

Data quality is kept with no phase change, no adverse effect on frequency distribution while keeping the true amplitude. The performance of the compression is more than fast enough to be usable during processing, it scales across CPU and GPU and utilizes any hardware to its full potential.

Counter intuitive to normal compression/decompression schemes, the Headwave compression/decompression is faster than accessing original data.

This enables work on much larger datasets in a much more efficient way, e.g. scanning through large volumes of seismic data for regional reconnaissance to dramatically reduce the time to generate prospects.

The 100% lossless compression reduces seismic acquisition data from half to 1/7th of the original size.

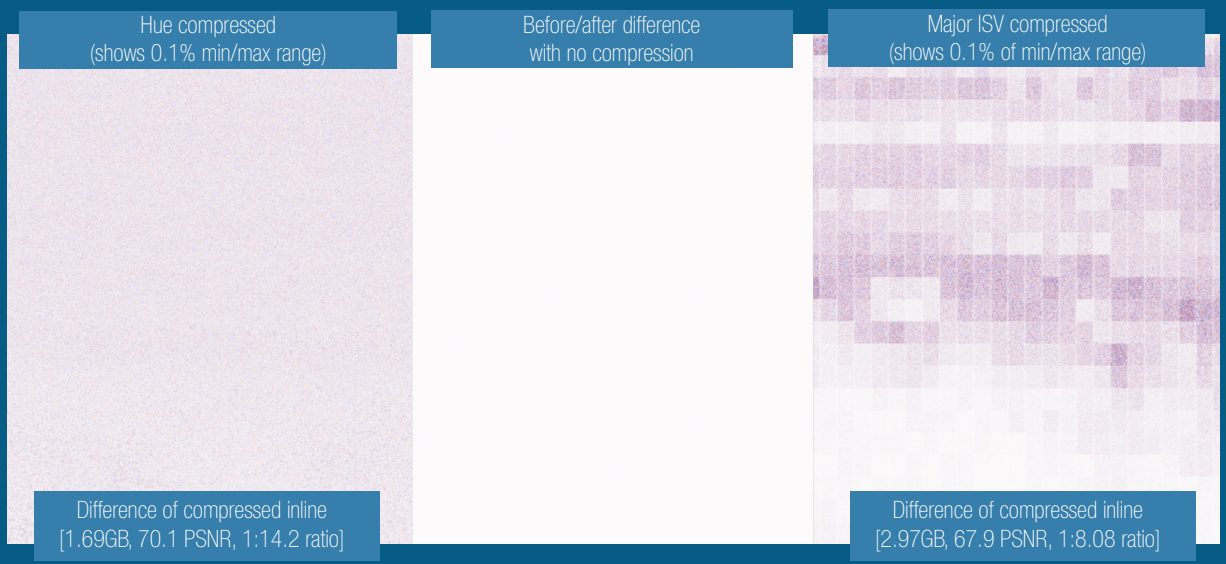
Compared to industry compression methods, this method performs better than all others at the same time this is done 8X to 15X faster. The algorithm is 1/6th to 1/14th more code efficient, resulting in a much better energy footprint.

For seismic acquisition companies, this means

- Acquire more data per node
- Faster data transfer off the node
- Positive energy balance
- Faster data transfer to customers

Real Life Quality Comparison - Hue Compression vs Leading Commercial Software

24GB North Sea Quad Dataset in SEG-Y format



	Commercial SW	Hue Compression	Improvement	
Resulting file size	2.97 GB	1.69 GB	1.76 X	50 % smaller files
Compression Ratio	8.08X	14.2X	1.76 X	
PSNR (Peak Signal to Noise Ratio)	67.8 dB	70.1 dB	2.3 dB	45 X faster bricking and compression
Compression speed	12.89 MB/sec	580 MB/sec	45 X	>100 X faster decompression
Completion Time from SEGY	45 min	1 min	45 X	

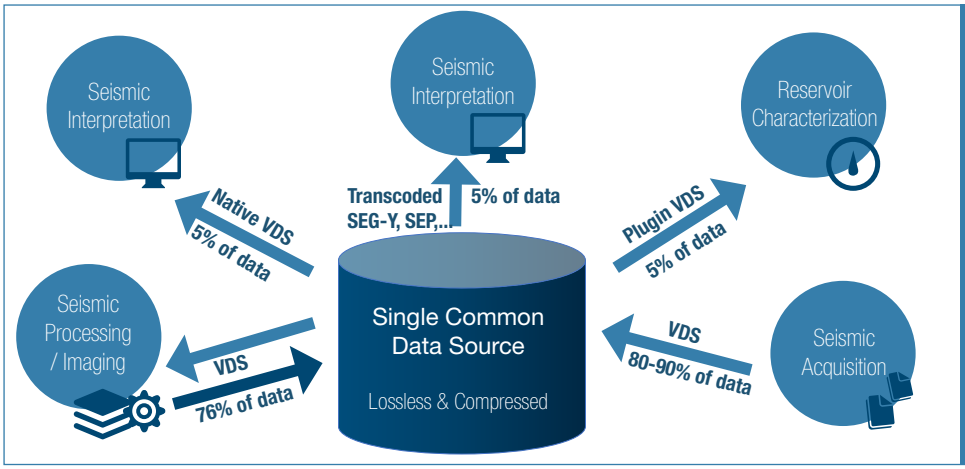


Figure: Application of compression

No data replication into different formats